|CURIEUX|

ACADEMIC REVIEW

January Issue

Issue 3 Part 3

Editing Staff

Chief-In-Editor

George Hasnah

Lead Editor

Amanuel Geremew

Ashley Du

Assisting Editors

Jing Yan Huang

Nidhi Potineni

Sophia Kong

Tony Yuan

Ridit Aggarwal

Manu Sharma

Rishi Gadikota

Eric Wang

Table Of Contents

Examining the Bases of Human Rights: A Philosophical Critique of Wolterstorff's Religious Grounding of Rights By William Oh

Abstract

Nicholas Wolterstorff, a philosopher and founder of reformed epistemology, proposes three essential conditions that any satisfactory account of inherent human rights must meet: it must (1) affirm that all humans possess dignity equally; (2) explain why human beings uniquely possess this kind of moral worth; and (3) make the violation of a right a moral injury to the rights-holder themselves, not merely an offense against cosmic balance, divine decree, or communal order. None of Islam, Buddhism, and Sikhism meets all of Wolterstorff's criteria. This paper conducts a comparative analysis of how each can be explored via Wolterstorff's criteria, finding that if one is willing to revise Wolterstorff's framework to emphasize agency, moral action, and relational sacredness, then Sikhism emerges as the most philosophically and ethically robust model for grounding human rights.

Introduction: Wolterstorff's Theistic Criteria for Rights

Nicholas Wolterstorff, in developing a theological foundation for human rights, argues that only a theistic grounding can provide a truly universal and morally robust notion of rights. Secular accounts based on autonomy, rational agency, or contractual consensus inevitably exclude those unable to meet these functional criteria—infants, people with profound cognitive disabilities, and the terminally ill, among others. These models—though historically influential—rely on capacities that not all humans possess, and therefore fail to uphold the idea of equal dignity for every human, which is essential to "human" rights. To address these limitations, in his original paper, Wolterstorff proposes three essential conditions that any satisfactory account of inherent human rights must meet: (1) it must affirm that all humans possess dignity equally, regardless of rational capacity or social utility; (2) it must explain why human beings uniquely—and not animals—possess this kind of moral worth; and (3) it must make the violation of a right a moral injury to the rights-holder themselves, not merely an offense against cosmic balance, divine decree, or communal order (Wolterstorff 266–67). Together, these criteria attempt to anchor rights in the person themselves, not in external validations or secondary functions.

According to Wolterstorff, only the Christian doctrine of God's unearned, unconditional love for every person can fulfill these demands (Wolterstorff 266). In this theological picture, value is not earned through ability or role, but is given and irrevocable, conferred by divine love. Yet, this raises an important question: Can other religious traditions offer alternative spiritual or moral foundations for human rights that satisfy the same philosophical rigor? Islam, Buddhism, and Sikhism each have coherent moral systems replete with views on human dignity, but their frameworks differ dramatically. This essay takes Wolterstorff's standard and applies it to each of these traditions.

I argue that while all three offer conceptions of moral worth, none ultimately meets all of Wolterstorff's criteria. Islam and Buddhism fall short in their own ways: Islam by emphasizing obligation over intrinsic moral standing, and Buddhism by rejecting permanent personhood and legalistic rights. Sikhism affirms dignity as both divine and actionable, which is the approach closest to that of Christianity. However, Sikhism also does not meet the second criterion by failing to clearly differentiate humans from other forms of life. Ironically, Christianity, which appears to meet the test on paper, ultimately collapses under the weight of its own assumptions when subjected to the critique of Paul Weithman—a Christian ethicist who argues that grounding dignity in divine love risks reducing humans to passive objects of God's affection, rather than active moral agents. If one is willing to revise Wolterstorff's framework to emphasize agency, moral action, and relational sacredness, then Sikhism—though not a perfect fit for Wolterstorff's criteria—goes beyond the framework and emerges as the most philosophically and ethically robust model for grounding human rights.

I. Islam: Divine Dignity Without Personal Rights

Islam offers a strong moral system through obligation, collective responsibility, and divine mandate. The Qur'an says, "We have certainly honored the children of Adam" (Qur'an 17.70) and affirms that all human beings possess dignity due to God's creative will. That dignity is not given on the basis of merit, reason, or task but by the simple fact of humanness. As Azra Awan explains, "human dignity is intrinsic in Islam, not granted by society but endowed by God at birth," and this God-given honor applies to all people equally, regardless of race, gender, or ability (Awan). Besides, the Qur'anic emphasis on equality—"Indeed, the most noble of you in the sight of God is the most righteous of you" (Qur'an 49.13)—is a theological commitment to racial, social, and gender equality. Here, Islam does succeed in meeting Wolterstorff's first condition: it believes that all human beings possess equal dignity, regardless of their capacity or social status.

In dealing with Wolterstorff's second test—human distinctiveness—Islam relies ultimately on *zameer*, or conscience. Human beings are credited with the capacity for moral judgment by most Islamic thinkers, and are therefore held accountable for their actions before God. The Day of Judgment itself is founded on such accountability: men and not animals are judged on their ability to distinguish between right and wrong. This moral capacity appears to distinguish man rationally from animals and gives Islamic anthropology justification for human uniqueness. M.S.H. Ma'sumi emphasizes that in Islam, "man is the only creature who possesses the power of moral discernment and accountability, making him the bearer of a special legal and moral status" (Ma'sumi 214). By basing the value of being good or bad in *zameer*, Islam seemingly meets Wolterstorff's requirement for human beings to possess a morally relevant quality not found in animals.

However, this model has another issue when applied to beings that lack moral consciousness. If moral value is bound up with *zameer*, then infants, the cognitively impaired, or potentially even the unconscious would lack basic dignity. Islamic theology has long been aware

of this problem. Traditional philosophers resolve it by appealing to *fitrah*, the innate nature toward good and truth that every human being is born with. From this viewpoint, even those who cannot exercise their conscience knowingly still participate in divine dignity through *fitrah*. As Awan notes, *fitrah* establishes that "every human being, by virtue of their creation, has inherent dignity and purpose," which includes even those who cannot articulate moral awareness (Awan). Islamic law also releases children and the mentally incapacitated from legal and moral culpability—not because they are worthless—but because their capabilities are not yet developed. Their rights and dignity remain with them, but their moral obligation is suspended (Traer 120).

Philosophically, however, this is difficult. If *zameer* is the virtue that morally distinguishes human beings from animals and some human beings are deficient in it, then the Islamic system risks excluding the very people that human rights are meant to defend. This model is susceptible to the same objection Wolterstorff brings against secular autonomy-based theories: that they base dignity on a capacity not shared by all humans. Therefore, while Islam does attest to universal dignity at the spiritual level, it cannot substantiate that dignity on a philosophically consistent basis that includes the non-autonomous. The use of *fitrah* is strong within the tradition, but does not explain why dignity should be granted to a human infant but not a sentient animal, given that both are unable to understand that something is right or wrong.

The third criterion—whether violations of rights are conceived as moral harms to the person—once more shows that the Islamic model has its limitations. Islamic ethics are predominantly duty-based. As Anis Ahmad indicates, "for every right, there is a corresponding duty," and the right to life means there is a corresponding set of obligations set forth for all living things (Ahmad 106). This model features mutual duties—between individuals and the community, and between human beings and God—but it does not describe rights as inherent possessions of an individual; rather, it expresses operations of divine law and social order that form them. Ma'sumi reinforces this by writing that "rights in Islam are embedded within a framework of duties to God and to society," and are not necessarily defined in terms of personal autonomy or subjective harm (Ma'sumi 217). Violations are typically seen as sins against God or against the shari'a, not necessarily as personal sins of the wrongdoer. Even when compensation mechanisms like diyya (blood money) exist, they are often framed as a means to restore social harmony rather than as acknowledgments of moral injury to the individual.

Over the course of the last few decades, reformers have pushed for a reformation of Islamic law to place more value on personal dignity and subjectivity, but these claims remain on the periphery of Islamic legal thinking. The orthodox tradition continues to be more about compliance with divine ordinance than about freedom for the rights-bearer. What this means is that disrespect for human dignity is generally viewed as disrespect for God's law and not disrespect for the person themselves (Traer 122).

In sum, while Islam succeeds in upholding universal human dignity in light of divine will and in distinguishing humans from animals based on moral capacity, it does not satisfy the third criterion, and when scrutinized more closely, the second of Wolterstorff's criteria is not satisfied

either. Its application of *zameer* as the basis for moral distinction fails in the face of non-autonomous humans, and its duty-based morality does not consider rights to be unbreakable moral possessions of individuals. Instead, rights are relative, situation-specific, and in a God-framework universe. Islam therefore shares the same problem Wolterstorff identifies in secular models: it cannot support dignity in the human person on the basis of themselves. It does secure obedience, loyalty to one another, and social duty—but not rights on the basis of the moral status of the individual.

II. Buddhism: Compassion Without the Rights-Holder

Buddhism has a completely different ethical framework than the Abrahamic religions—one rooted not in divine mandate or intrinsic value, but rather in unity, the path to enlightenment, loving-kindness, and the elimination of suffering. Its vocabulary of morals is built around core doctrines such as the Four Noble Truths, the Eightfold Path, and *paṭicca-samuppāda* (dependent origination), all of which collectively emphasize the interdependence of sentient beings and that every action has its reward in the wheel of *samsāra* (Keown 4). From the very beginning, Buddhist tradition has held empathy, restraint, and mindfulness as the foundation of ethical living.

This system fulfills Wolterstorff's first requirement most powerfully since Buddhism places the suffering of all sentient beings at the center of its moral framework, not only including infants and cognitively disabled individuals but also animals and even lower forms of life within the range of its concern. Ethical worth is not based on rationality or autonomy, but on the capacity to suffer. Compassion (*karuṇā*) and loving-kindness (*mettā*) are bestowed on any creature capable of experiencing suffering (Keown 7). The bodhisattva ideal—the wise one chooses to remain in the cycle of suffering so he can assist others—is a paradigm of radical ethical inclusiveness (Keown 10-12). In this, Buddhism perhaps surpasses Wolterstorff's own anthropocentric explanation in extending moral concern outside the human species.

However, this very openness presents problems for Wolterstorff's second and third criteria. Buddhism is not strict on the ontological line between humans and animals. The doctrine of *anattā* (no-self) challenges the idea of an abiding, persistent person who would be able to function as a rights-bearer in the classic Western conception (Keown 5). Instead of a morally unique human soul or essence, there are only dynamic aggregations of causes and conditions. Sentient beings are all part of the same karmic flux. As a result, the second requirement—that moral worth to be uniquely and exclusively human—cannot be satisfied. Buddhism explicitly denies the moral and philosophical hierarchy that Wolterstorff demands.

The third criterion also proves deeply problematic. Buddhism does not have a legal or theological conception of violations of rights as individual moral wrongs. Wrongdoing is embodied in terms of karma: an act produces consequences for the actor based on intention and attachment, but not necessarily on the effect it has on the harmed party. In Buddhist thought, harm is more an issue of spiritual blockage than violation of another's dignity. Since there is no enduring self to be harmed, there can be no such thing as rights understood as unbreachable

claims. Keown has attempted to reconstruct Buddhist ethics in the language of relational rights, but this requires retrofitting a Western model of rights to a tradition that simply does not encompass stable persons with inherent moral rights (Keown 15–18).

Even righteous moral acts such as a *bodhisattva* sacrificing their life for others are not thought of as the exercise of someone else's right to life or protection. Rather, they are expressions of boundless compassion for beings trapped in suffering. Buddhism is about intention, morality, and spiritual growth, and not justice or compensation. It is about minimizing suffering, not rights enforcement. Thus, Buddhism also fails to pass Wolterstorff's second and third tests: it does not clearly distinguish humans as the sole or primary bearers of rights, nor does it frame injustice as moral harm to the self. It is steeped in ethical depth and universal compassion but lacks the ontological clarity and legal specificity required for a robust modern theory of rights.

III. Sikhism: Divine Light Without Species Distinction

Sikhism offers a singularly powerful theological and ethical foundation for human dignity by basing value in the teachings of *jot*—the divine light said to be present within every living thing as equally. As formulated in Sikh scripture and tradition, all human beings irrespective of origin, capacity, or condition, have this spark of divinity. Amarpreet Singh Munde explains that "everyone is an embodiment of the same divine light, and therefore no one shall be less than the other" (Munde). This metaphysical justification of sacred equality directly meets the first of Wolterstorff's three conditions. Dignity in Sikhism is not bestowed by social function, intelligence, or divine whim, but is rooted in the very fabric of existence itself. All human beings are sacred because all human beings are divine at their very core.

Sikhism also places a particular emphasis on moral action and not on sheer metaphysical agreement. Through *seva* (selfless service), *gurmat* (moral discrimination under the scrutiny of divine wisdom), and the two-fold concept of *Miri-Piri* (the union of religious and political power), Sikhism cultivates an ethical life centered on duty and justice (Singh). The faith does not ask its followers merely to believe in the other person's dignity—it asks them to act on it. Guru Tegh Bahadur's life, which he sacrificed for the sake of the religious freedom of non-Sikhs, is a living demonstration of this morality (Singh). Sikhism thereby integrates moral agency into spiritual identity. This ontological relationship of living with moral responsibility satisfies Wolterstorff's first requirement in both form and deed: it describes universal, unearned dignity and then requires action in the world based on that reality.

But when we turn to Wolterstorff's second requirement—the requirement that this moral value be unique to human beings—Sikhism hits a roadblock. Because divine light is said to be inherent in all manifestations of life, Sikh metaphysics obscure the categorical distinction between human and animal. While humans are specifically commanded to respond to their divine light—to help others, to strive towards justice, to live ethically—Sikh theology does not bestow sanctity on the human alone. This broad ontology is a challenge to Wolterstorff's demand that rights must be grounded on what makes humanity different from animals. Humanity is not

the sole possessor of moral value in Sikhism. Wolterstorff does not simply ask that humans be morally differentiated from animals; he asks for unique moral characteristics—a difference in kind, not degree.

This might appear to be a deviance, but it also assumes an expansive moral horizon. Where Wolterstorff wants to affirm human uniqueness for the sake of protecting the integrity of rights, Sikhism tries to extend moral regard without abandoning human agency. In other words, Sikhism avoids Christian theology's human exceptionalism as well as utilitarian moral leveling; but because it offers no metaphysical account of uniquely human sacredness, it cannot meet Wolterstorff's second condition. Sikhism affirms human distinctness in responsibility rather than essence.

Regarding the third requirement—that violations of rights need to be imagined as moral hurts to the person—Sikhism surpasses all other traditions discussed here. It imagines injustice as violating the divine light in the person. Harms are not only cosmic imbalances or theological transgressions against God, but attacks on the sacred dignity in the person. This is not a piece of abstract theology but a firm principle in Sikh history. The Khalsa, founded by Guru Gobind Singh, along with the protection of religious minorities and the martyrdom of Sikh heroes, expresses a theology in which harm to one is seen as harm to all (Singh). Sikhism thus fulfills the third requirement better than Christianity, Islam, or Buddhism.

However, it should be noted that Sikhism does not completely pass Wolterstorff's test. In maintaining the dignity of all creatures, it cannot offer a strictly anthropocentric rights-based theory. Its metaphysical turn, while rich in ethics, diverges from the narrowly defined philosophical framework Wolterstorff puts in place. However, in its integration of dignity, action, and justice, it begins to diverge outside the framework altogether. Sikhism grounds rights not merely in being beloved (as Christianity), or in moral order (as Islam), or in pity (as Buddhism), but in ongoing activity of beholding and upholding the sacred presence in others. Its distance from a model grounded in exclusivity is not a failure but a sign of its philosophical reach beyond it.

IV. Christianity's Internal Collapse: Weithman, Dudrick, and Singer

It is important to emphasize that this philosophical critique of Wolterstorff's metaphysical model should not be taken as a dismissal of Christianity's historical role in advancing human rights. From William Wilberforce's abolitionist campaigns to Martin Luther King Jr.'s vision of non-violent resistance, and even earlier in the Christian imperial reforms of Theodosius I—who outlawed forced prostitution—Christian belief has fueled major movements for justice. The term "social justice" itself originates from the Papal Encyclical Rerum Novarum (1891), demonstrating the tradition's long-standing concern for moral and civic reform. This essay does not dispute Christianity's practical efficacy in inspiring human rights activism. Rather, it argues that Wolterstorff's specific theological framework—when evaluated against his own philosophical criteria—struggles to provide a consistent account of rights as grounded in the moral status of the person. Sikhism, by contrast, embeds both sacredness and agency within the

self and commands action not just through divine love, but through shared moral responsibility. It is on that conceptual basis—not historical effectiveness—that this paper makes its comparative claim.

Wolterstorff's own Christian account is, at least on the surface, the only tradition to meet all three of his demands. Building human rights on the doctrine of the imago Dei—the doctrine that all human persons are created in the image of God—Christianity teaches that every human person is utterly and without reserve loved and prized by God. This divine love, according to Wolterstorff, confers intrinsic worth, regardless of the intellectual capacity, social worth, or moral standing of a person. Following from this interpretation, rights are "what is owed a person because of their worth," and worth is bestowed by God's attachment love (Attridge 213). Because this kind of love is permanent and non-merit-based, Christianity appears to pass the first test.

The second requirement is also satisfied in form. Christian theology sharply distinguishes between human and animal. Only human beings are created in the image of God and given souls, and can respond to divine love and command. Animals are respected in the majority of Christian traditions but not as rights-holders in the moral or theological sense. Christian anthropology thus makes the human ontologically and morally distinct, satisfying the requirement of exclusivity.

While the first two conditions present manageable challenges, the situation becomes far more complex when we consider the third: that violation of a right must constitute an offense to the moral worth of the person. It is here that Wolterstorff's model comes unstuck. Given that human dignity is not inherent but conferred—it depends on God's love and not on human effort—the model starts to externalize moral worth. Critics like Paul Weithman have added to this argument, arguing that if human persons are valuable because they are loved by God, then their rights are more offensive to God when they are violated than they are to the person. In his provocative essay "God's Velveteen Rabbit," Weithman compares Wolterstorff's human subject to a child's beloved doll: loved, indeed—yet passive, silent, and subject to the pleasure of another (Weithman 408). The result is a model in which human beings are sacred property rather than independent moral agents.

This is supplemented by David Dudrick's "afterlife argument," a challenge to what it is to ground rights in transcendent destinies. If the individual's moral worth is to be grounded in their position with God in the afterlife, then present injustices may be minimized or deferred. "Such a person may not now but—according to the theist's belief in an afterlife—one day will exercise the capacities of her nature. She is, therefore, of great worth precisely as a bearer of such a nature" (Dudrick 14). The risk is that human rights are now theological vows, rather than immediate ethical obligations. "In that case, it's doubtful that Wolterstorff would tell his neighbor to junk the Jaguar; indeed, he likely would counsel him to keep it under a dust cover and to lift the cover periodically to admire it, attempting to see in it what the mechanic will see" (Dudrick 14). Dudrick's concern is that this model postpones moral seriousness and the immediacy of moral injury. Injustice now becomes tolerable because it is framed as temporary or

redemptive. "That, I suggest, is a better analogy for the theist's take on [a] situation involving the severely disabled person" (Dudrick 14).

Peter Singer, despite being as much a utilitarian as he is, offers insightful contrast. While he reduces moral worth to interest and sentience, he places suffering and moral damage firmly in the lived experience of the subject. "The capacity for suffering and/or enjoyment or happiness is not just another characteristic like the capacity for language… it is a prerequisite for having interests at all" (Singer). Singer may not succeed in distinguishing human persons from animals as Wolterstorff does, but his ethic ensures that harm matters because of what it does to the one harmed—rather than because of what it reveals about divine love or metaphysical intention. "If a being suffers, there can be no moral justification for refusing to take that suffering into consideration…the principle of equality requires that its suffering be counted equally" (Singer). This emphasis on immediate experience foregrounds the actual injury to the person rather than the theological interpretation of that injury. Singer critiques theological abstractions by noting that "talk of intrinsic dignity or moral worth only takes the problem back one step… fine phrases are the last resource of those who have run out of arguments" (Singer).

Wolterstorff attempts to salvage his account by suggesting a dynamic reading of the imago Dei, which states, so far as people are loved, they are commissioned to do things in God's name (Wolsterstorff 268). This rescue, however, does not answer Weithman's or Dudrick's concerns. Value is still motionless in God's love at the center of the model, and not in autonomous moral agency. Therefore, Christianity is deficient exactly where it appears to come through. Its emphasis on value received positions the individual in a state of dependence, not independence. Far from being the highest example of human rights, Christianity—using Wolterstorff's own standards—exposes the weakness of any model that does not center on the rights-holder.

V. Sikhism Revisited: A Better Model Than the Criteria

If no tradition—least of all Christianity—can fully satisfy Wolterstorff's model for the ground of basic human rights, then maybe the model itself needs to be rejected. Wolterstorff's demands prefer a model that emphasizes value on divine gift, clearly distinguishes humans from all other beings, and interprets moral harm as something felt subjectively and strongly by the rights-bearer. Is it possible, then, that the most powerful ethical intuitions at the heart of these standards—those concerning pain, individuality, and worth—can survive even without Wolterstorff's metaphysical boundaries?

When we return to Sikhism with this in mind, we might recognize that while it technically falls short in one of the standards, it may actually succeed in an even deeper sense. Sikhism fails to meet the second requirement in that it rejects the idea that human beings alone are sacred. Its *jot* doctrine does claim the presence of divine light in all life forms, including animals, plants, and the cosmos (Munde). Still, this refusal to portray human beings as uniquely valuable does not dissolve moral differences or render human rights obscure. Instead, Sikhism preserves a special moral calling for human beings. Only humans are morally responsible for

realizing divine light and responding to it. Only humans can do *seva*, discern *gurmat*, and embody the *Miri-Piri* principle (Singh). Thus, while Sikhism denies the singularity of divine presence, it affirms the singularity of moral responsibility. Sacred is everywhere, but responsibility is uniquely human.

In addition, Sikhism enhances our understanding of moral injury. For Sikh philosophy, injustice is not merely a violation of holy law or breakage of social order—it is a desecration of the holy within (Munde). When someone is harmed, his or her dignity is not lost socially or symbolically; his or her inward sanctity is violated. This understanding of harm is closer to Wolterstorff's third requirement than any of the other previously discussed traditions. In fact, it expands that criterion by establishing the sense of moral harm not so much in subjective suffering or legal right, but in metaphysical desecration. The affront is not so much against one's right—it is an affront to the sacred.

Sikhism also elevates Singer's utilitarian paradigm. While Singer rightly puts stress on the capacity to suffer and onto moral exigency, his approach reduces moral status to a calculable function. Sikhism resists this reduction; it offers a picture of moral concern which is metaphysically thick, non-hierarchical, and non-instrumental. Unlike Singer, who builds ethics from below (from experience), Sikhism builds from above (from the sacred). While it differs from Wolterstorff, it does not place moral value beyond the individual. It internalizes sanctity and calls each person to be mindful of it and to respond to it.

Thus, even if Sikhism fails Wolterstorff's technical criterion of exclusivity, it is a more unified, comprehensive, and action-guiding human rights theory. It reframes rights as sacred, presence-based moral obligations, that are instantiated in lived protest against injustice.

Conclusion: Holy Obligation in a Common World

This essay began by taking Nicholas Wolterstorff's argument that any adequate explanation of human rights will satisfy three conditions: it will claim dignity for all, distinguish humans as especially worthy, and categorize rights violations as moral wrongs to the victimized individual. Upon application of these criteria, Islam, Buddhism, and Sikhism all fall short at first. Islam focuses on divine command and civic responsibility, but ultimately equates moral harm with disobedience to God instead of harm to the self. Buddhism, though rooted in profound compassion and ethical interdependence, does not frame humans as exclusive rights-bearers—its moral universe encompasses all sentient beings equally. Moreover, its denial of a permanent self (anattā) undermines the very idea of a rights-holder capable of suffering moral injury in the individualistic sense. There is moral harm, but it is distributed across karmic networks rather than localized in the person. Sikhism affirms sacred worth and moral responsibility, yet spreads divine light to all sentient life, and thereby appears to lack the necessity of exclusivity. Christianity, the faith of Wolterstorff, seems to pass all three tests—until it is subject to critical examination. Paul Weithman, David Dudrick, and Peter Singer each uncover aspects in which the Christian approach makes the human subject passive, delays moral necessity, or outsources value into divine love. The person is made lovable, yet may not be culpable. The harm gets

translated into theology, but maybe not into individual harm. In these criticisms, even Christianity fails under the full weight of its own test.

Sikhism rethought and reconsidered can offer not only a different perspective but an improvement. It meets the first and third criteria with moral and metaphysical richness, and though it does not meet the second criterion's demand for anthropocentrism, it offers something deeper than that: a vision of universal sacredness along with uniquely human responsibility. Its theory of rights is based not on superiority, but on vocation. Its response to injustice is not to call upon divine protection, but to take on divine responsibility. Sikhism therefore transforms rights from static inheritances into sacred duties. It bridges ontology with action, theology with ethics, and individual dignity with collective struggle.

In a time of global crisis, when human rights are invoked as both a mantle and a slogan, we need a base that is both spiritually powerful and philosophically resilient. Sikhism offers such a model. Sikhism not only meets Wolterstorff's requirements—it reinterprets them. It reminds us that dignity is not only conferred but enacted, that harm is not merely endured but actively countered, and that rights are more than assertions—they are sacred obligations we owe one another in the sight of the divine.

Works Cited

Ahmad, Anis. "Human Rights: An Islamic Perspective." Policy Perspectives, vol. 3, no. 1, 2006, pp. 101–112. JSTOR, www.jstor.org/stable/42909260.

Attridge, Harold W. "Wolterstorff, Rights, Wrongs, and the Bible." The Journal of Religious Ethics, vol. 37, no. 2, 2009, pp. 209–219. JSTOR, www.jstor.org/stable/25654517.

Awan, Azra. "Human Rights in Islam." WhyIslam.org, 3 Apr. 2025, www.whyislam.org/human-rights-in-islam.

Dudrick, David. Why Tolerate Religion? A (Surprising) Nietzschean Answer. Center for Philosophic Exchange, hdl.handle.net/20.500.12648/7675. Accessed 27 July 2025.

Keown, Damien. "Are There 'Human Rights' in Buddhism?" Journal of Buddhist Ethics, vol. 14, 2007, pp. 1–27. blogs.dickinson.edu/buddhistethics/files/2010/04/keown.pdf.

Ma'sumi, M. S. H. "Islamic Concept of Human Rights." Islamic Studies, vol. 11, no. 3, 1972, pp. 211–221. JSTOR, www.jstor.org/stable/20846914.

Munde, Amarpreet Singh. "Sikhism & Human Rights." Sikh Missionary Society UK, Accessed 27 June 2025, www.sikhmissionarysociety.org/sms/smsarticles/sikhismhumanrights.

Singer, Peter. "All Animals Are Equal." Animal Rights and Human Obligations, edited by Tom Regan and Peter Singer, Prentice Hall, 1989, pp. 198.

Singh, Gurmukh. "Sikhism and Human Rights." Sikh Missionary Society UK, Accessed 27 June 2025, www.sikhmissionarysociety.org/sms/smsarticles/sikhismandhumanrights.

Traer, Robert. "Human Rights in Islam." Islamic Studies, vol. 28, no. 2, 1989, pp. 117–129. JSTOR, www.jstor.org/stable/20840066.

Weithman, Paul. "God's Velveteen Rabbit." The Journal of Religious Ethics, vol. 37, no. 2, 2009, pp. 393–410. JSTOR, www.jstor.org/stable/25654523.

Wolterstorff, Nicholas. "Justice as Inherent Rights: A Response to My Commentators." The Journal of Religious Ethics, vol. 37, no. 2, 2009, pp. 261–279. JSTOR, www.jstor.org/stable/25654520.

Mapping Personality to Music: A Review of Big Five Influences and Applications
By Advika Kamath

Abstract

     Research on the impact of the Big Five personality traits on musical preferences has produced inconsistent results, largely due to variations in methodology and interpretation. This review aims to address these inconsistencies by drawing on the diverse approaches and preferences measured within music used across studies and relating them to one another to establish a more conclusive understanding of the relationship between personality traits and musical preferences. Specifically, I have structured the analysis to show how each of the five personality traits influences music preferences, drawing on findings from various studies that used different methods to explore these relationships. My analysis shows that some personality traits such as openness and extraversion are strong predictors of music preferences. The trait Openness to experience reflects a cluster of individual differences that center on receptivity to new ideas and experiences (McCrae and Costa). In line with its label, openness to experience is associated with a preference for specific music styles. In contrast, while the traits Conscientiousness and Agreeableness had weak effects on music preferences, they consistently correlated with the characteristics of their respective personalities. However, the impact of Neuroticism on personality traits has varied across previous studies, resulting in inconsistent findings. Each of the traits provided findings of musical preferences that reflected characteristics of those traits with some variations. Together, these steps clarify when and how personality meaningfully informs music choice and how to apply it responsibly.

Introduction

     Personality psychology studies how personality develops and how traits shape behavior, helping build self-awareness, strengthen relationships, predict behavior, and support well-being (Cherry). The "Big Five" traits of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism provide a widely used framework for describing patterns of feeling, thinking, and behavior. Briefly: Openness captures curiosity and imagination; Conscientiousness reflects organization and goal-directedness; Extraversion involves sociability and positive affect; Agreeableness reflects prosocial tendencies; and Neuroticism reflects emotional instability (Cherry).

     The Big Five model of personality can be used to explain individual differences in music preferences, as people often choose music that reflects their traits, and those traits can, in turn, bias what they are drawn to (Cherry; Liu). For example, individuals high in openness tend to enjoy more complex or intense styles (e.g., classical, some rock), and more outgoing, cheerful listeners often prefer upbeat, conventional music (e.g., pop, country) (Cherry). Several psychological accounts explain why traits align with musical tastes. The Uses and Gratifications Theory holds that people actively choose media, including music, to satisfy needs such as identity, social connection, mood regulation, and tension release (Vinney). Arousal Theory

suggests listeners seek music that helps maintain optimal arousal (e.g., higher-arousal music for stimulation, lower-arousal music to feel calm) (Nickerson). Empathizing–Systemizing Theory further proposes that cognitive styles shape preferences for emotionally rich vs. highly structured music (Goally). Together, these perspectives help clarify how personality can guide music choices, and empirical work supports the general link between personality traits and music preferences (Cherry; Liu).

   Previous research has explored the relationship between personality traits and music preferences, but findings have often been inconsistent due to variations in methodology and measurement standards. This literature review aims to address these inconsistencies by examining the range of approaches used to define and assess musical preferences, offering a streamlined synthesis to support a more unified understanding of this complex relationship. Focusing on the Big Five personality traits, the review places particular emphasis on genre-specific patterns among young adults. By integrating psychological and sociological perspectives, it investigates how traits such as openness to experience and extraversion are associated with preferences for specific musical genres, dimensions, and characteristics. Additionally, the review highlights key methodological and theoretical gaps, laying the groundwork for future research to further clarify the links between personality and musical taste in emerging adult populations.

Theory

<u>STOMP and MUSIC Models</u>

   Prior research on music preferences has used a variety of models to describe and measure music preferences. The Short Test of Musical Preferences (STOMP) model is a 14-item scale assessing preferences in music genres by assessing participants' liking or disliking on a self-report scale (Rentfrow and Gosling). There are four broad musical dimensions that are examined: Reflective & Complex, reflecting music that is relatively slower in tempo, uses mostly acoustic instruments, and has very little singing (i.e., blues, jazz, classical, and folk genres); Intense & Rebellious, which includes music that is faster in tempo and uses mostly electric instruments (i.e., rock, alternative, and heavy metal genres); Upbeat & Conventional, which includes music that is moderate in tempo and uses both acoustic and electric instruments (i.e., country, pop, religious, and soundtrack genres); and Energetic & Rhythmic, reflecting music that is moderate in tempo and uses primarily electric instruments (i.e., rap/hip-hop, soul/funk, and electronic/dance genres).

   Another widely accepted framework for conceptualizing musical preferences is the MUSIC model, which is a five-factor structure consisting of five genre-free dimensions: Mellow, Unpretentious, Sophisticated, Intense, and Contemporary. The Mellow factor features romantic, slow, and quiet attributes as heard in soft rock, R&B, and adult contemporary genres. The Unpretentious factor includes uncomplicated, relaxing, and unaggressive attributes as heard in country genres. The Sophisticated factor is characterized by a preference for complex musical

forms like classical, opera, jazz, and world music. The Intense factor includes distorted, loud, and aggressive attributes as heard in classic rock, punk, heavy metal, and power pop genres. Finally, the Contemporary factor features rhythmic, upbeat, and electronic attributes as heard in the rap, electronica, Latin, and Euro-pop genres (Rentfrow, Goldberg, and Levitin).

 Acoustic features

Other studies of music preferences seek to qualify and measure acoustic features of music, including dynamics, type of piece, mode, register, and tempo. Dynamics is the variation in loudness between notes or phrases. By indicating how loud or soft music should be played, dynamics shape the emotional impact of a song and influence listeners' music preferences. Building on structural flexibility in music, modes are scale patterns formed from notes within a scale. Modifying these modes reveals which tonal characteristics individuals favor, whether bright and uplifting or dark and tense (Young, Dickman, and Ogn). Another element related to structure is register, which is the range of pitches an instrument or voice can produce. Registers offer insight into individuals' music preferences by creating different moods and effects in music. This helps us understand the contexts in which people listen to music, depending on the register involved. In addition, an important structural factor is tempo, which refers to the speed or pace at which a piece of music is played. Tempo dictates how fast or slow the rhythm and notes of a song unfold, which shapes how individuals feel, interpret, and perform music. This aspect is why tempo often defines the character of a musical genre and determines the mood and emotion of the piece (Music Pandit).

Musical features

Musical features describe different aspects of songs' sound qualities and how they may be perceived by listeners. These include acousticness, danceability, energy, instrumentalness, liveness, tempo, and valence, and they are defined according to the Spotify Web API (Moreira Júnior et al.). These features are fundamental to determining the music preferences of individuals as they capture objective elements such as tempo and acousticness along with perceived qualities like valence and energy.

Acousticness describes the degree to which the music is acoustic or electronic in nature. This connects closely with danceability, which describes how suitable a track is for dancing, based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity (Moreira Júnior et al.). Danceability is further shaped by energy, which represents a perceptual measure of intensity and activity. Another important feature is instrumentalness, which predicts whether a track contains no vocals or only non-lyrical vocal sounds. Alongside this, liveness detects the presence of an audience in the recording, providing information about the performance setting. Tempo also plays a central role, referring to the overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration (Moreira Júnior et al.). Finally, valence indicates a track's musical positiveness. High valence means the

track sounds positive, such as happy, cheerful, or euphoric, while low valence means it sounds negative, such as sad, depressed, or angry.

Openness

Research consistently highlights Openness to Experience as a key predictor of music preferences, showing a strong positive association with the Reflective & Complex dimension and a moderate positive association with the Intense & Rebellious dimension, alongside a negative association with Upbeat & Conventional music. Studies consistently demonstrate that individuals high in Openness tend to prefer music that is novel, emotionally intense, and cognitively stimulating, while showing less interest in mainstream or formulaic genres (Rentfrow and Gosling; Vella and Mills; Langmeyer et al.; Moreira Júnior et al.; Schäfer and Mehlhorn). While these studies converge on the link between Openness and preferences for non-mainstream musical styles, they differ in methodological depth. For instance, Langmeyer et al. confirmed the predictive value of Openness but did not consider contextual influences such as musical function, listener expertise, or perceptual qualities like tempo and loudness. Similarly, Rentfrow and Gosling—though foundational in establishing the association between personality traits and musical preferences—acknowledged a key limitation: they recommended that future research explore the relationship between cognitive complexity of listeners and the complexity of musical structure, which their STOMP model did not adequately capture. This limitation was directly addressed by Vella and Mills, who incorporated not only personality traits but also how individuals use music cognitively and emotionally. They found that the link between Openness and musical preferences is functionally mediated, such that cognitive uses of music explained preferences for Reflective & Complex styles, while emotional uses partially mediated preferences for Intense & Rebellious styles, adding a nuanced layer to the personality–preference relationship.

However, all three studies relied on the STOMP model, which categorizes music using a fixed genre-based structure. This limits cross-study comparability due to inconsistent or evolving genre boundaries. In response, Moreira Júnior et al. used a data-driven method based on Spotify audio features (e.g., valence, tempo, danceability) and personality data extracted from users' Facebook profiles. Their findings further clarified that Openness was negatively correlated with valence, suggesting a preference for emotionally complex or melancholic music. They also observed negative correlations between Openness and both mainstream and danceable genres, highlighting a clear affinity for music that is structurally intricate, introspective, and less geared toward rhythmic simplicity or social immediacy. Additionally, Flannery and Woolhouse found that, although there was no main effect of Openness, there was a link between dynamics, piano dynamics, and the type of piece played to Openness.

A more comprehensive and standardized solution comes from Schäfer and Mehlhorn's meta-analysis, which employed the MUSIC model. Their findings reinforced earlier conclusions: Openness was most strongly associated with Mellow, Sophisticated music, with smaller positive correlations with Intense styles, offering a more generalizable and consistent framework for

cross-study comparison. The results additionally showed that individuals who scored higher on Openness appeared to like music more in general. These people seemed to be open to all kinds of music, as all mean correlations were positive.

Together, these studies suggest that Openness to Experience is reliably linked to preferences for music that is aesthetically rich, intellectually engaging, and emotionally nuanced, rather than conventional or commercially driven. This suggests that Openness may involve greater emotional and aesthetic receptivity, leading to broader musical engagement overall. Furthermore, its broad positive associations across all music dimensions suggest that it may reflect a general openness to musical experience rather than specific genre preferences.

Conscientiousness

Conscientiousness has consistently emerged as one of the weaker predictors of music preference within the Big Five personality traits. Although Langmeyer et al. found small bivariate correlations, showing a negative association with Intense & Rebellious music and a weak positive association with Upbeat & Conventional music, their structural equation modeling (SEM) revealed no significant unique effects when controlling for other traits. This suggests that Conscientiousness does not independently influence musical taste once stronger predictors, such as Openness and Extraversion, are accounted for. Although Moreira Júnior et al. found no broad effects of Conscientiousness on most musical features, they did observe a modest positive correlation with tempo. Similarly, Flannery and Woolhouse found no significant main effect of Conscientiousness, but there were significant interactions of Conscientiousness with mode, piece, and tempo. Individuals with low Conscientiousness rate slow tempo as the lowest, but there is no significant difference for fast tempo. These results demonstrate how different levels of Conscientiousness influence preferences for music type, mode, and tempo.

Separately, Schäfer and Mehlhorn's meta-analysis reported a small, consistent negative association between Conscientiousness and Intense music, suggesting that more conscientious listeners are slightly less drawn to rebellious, aggressive, high-arousal styles. Rentfrow and Gosling also observed this pattern, noting a small positive correlation with Upbeat & Conventional music and a negative correlation with Intense & Rebellious music.

These findings are consistent with the psychological profile of highly conscientious individuals, who typically value order, control, and structure. Upbeat & Conventional music, which is often defined by clear melodies, familiar structures, and socially accepted themes, aligns with these preferences. In contrast, intense music, which often features unpredictable dynamics, loud instrumentation, and themes of nonconformity, may be perceived as overstimulating or chaotic. This contrast helps explain the weak but consistent pattern of conscientious individuals favoring conventional genres while avoiding more intense, rebellious styles.

Extraversion

Extraversion has consistently emerged as the second strongest predictor of musical preferences after Openness, as demonstrated in foundational and meta-analytic studies (Rentfrow and Gosling; Schäfer and Mehlhorn). Across studies, Extraversion has been positively associated with the Energetic & Rhythmic and Upbeat & Conventional dimensions, reflecting music that is lively, danceable, and socially engaging, such as pop, hip-hop, and electronic music.

For example, Vella and Mills and Rentfrow and Gosling found that Extraversion significantly predicted preferences for both Energetic & Rhythmic and Upbeat & Conventional styles. These genres, often featured in social or party settings, resonate with the extraverted individual's desire for stimulation, emotional expressiveness, and social connection. Although Vella and Mills hypothesized that background use of music would mediate the relationship between Extraversion and genre preference, their analysis found no such mediation, suggesting a more direct link between trait disposition and musical taste.

Expanding on this, Langmeyer et al. identified a moderate positive correlation between Extraversion and Intense & Rebellious music preferences, indicating that extraverts may also gravitate toward emotionally intense and high-arousal genres, not just those associated with mainstream or dance settings. Similarly, Moreira Júnior et al. found that Extraversion was associated with a preference for danceable music and popular tracks, further reinforcing the pattern that extraverts are drawn to stimulating and accessible musical experiences.

These findings are corroborated and refined by Schäfer and Mehlhorn's meta-analysis, which employed the MUSIC model. Their analysis showed a weak but consistent positive correlation between Extraversion and Contemporary music, a category encompassing modern, rhythmic, and socially prominent genres. While the strength of this association is smaller compared to Openness–Sophisticated music links, it nonetheless confirms the trait's relevance across methodologies and classification models. Flannery and Woolhouse found no main effect of Extraversion, but small interactions were found: high-extraversion listeners preferred major mode and high register, while low-extraversion listeners rated forte lowest and disliked slow tempo. When it came down to piano dynamics, individuals higher in Extraversion rated it lower compared to individuals high in Openness.

Taken together, these studies form a coherent narrative: highly extraverted individuals—characterized by sociability, energy, and a preference for external stimulation—consistently show a preference for music that is upbeat, rhythmic, emotionally charged, and culturally popular. Their musical taste appears to reflect their broader behavioral tendencies, favoring genres that match their outgoing, expressive, and energetic personality. In contrast, individuals who were lower in Extraversion preferred softer, but medium-fast-paced music with less dynamic range, indicating that individuals higher and lower in Extraversion do not display opposing music preferences. Although they preferred softer music, they still exhibit similarities with those high in Extraversion due to their preference for music with a higher tempo.

Agreeableness

Research findings characterizing the relationship between Agreeableness and music preferences have been similar but have differed in magnitudes across studies. Langmeyer et al. found that Agreeableness was positively associated only with the Upbeat & Conventional music dimension. Similarly, Moreira Júnior et al. found that Agreeableness was associated with a stronger preference for music high in danceability. Supporting this trend, Schäfer and Mehlhorn's meta-analysis revealed a small but consistent link between Agreeableness and a preference for Unpretentious music. Rentfrow and Gosling also reported positive associations between Agreeableness and both the Upbeat & Conventional and Energetic & Rhythmic dimensions, further reinforcing the findings of Langmeyer et al. Flannery and Woolhouse found a main effect of Agreeableness, with highly agreeable participants rating the modified musical excerpts (excerpts with register, mode, and tempo variations) higher overall. In addition, there were several significant interactions. For piece and Agreeableness, Bach excerpts were rated disproportionately higher by participants high on Agreeableness. Slow tempo was disproportionately less preferred for low Agreeableness.

These findings depict that positive, friendly, and energetic music resonates with agreeable individuals' preference for pleasant, uplifting, and socially comforting environments. Agreeable people may be drawn to these music genres because of their upbeat, familiar, and harmonious qualities that reflect their own social and emotional traits centered around friendliness and cooperation.

Neuroticism

Findings for Neuroticism have been inconsistent across studies. Vella and Mills found that trait Neuroticism correlated positively with emotional uses of music and marginally with background uses. This supported their hypothesis that people high in trait Neuroticism are more likely to use music for emotion regulation, particularly during emotional strain, as they often report greater stress and depressive symptoms. The authors also expected these individuals, given their heightened sensitivity, to avoid emotionally intense or aggressive music. However, trait Neuroticism was not significantly related to preferences for Intense & Rebellious music, contrary to expectations.

Expanding on these patterns, Langmeyer et al. similarly reported only a weak negative effect of Neuroticism on preference for Intense & Rebellious music, suggesting a mild tendency to dislike this music style. In contrast, they found that Neuroticism had a strong positive effect on preference for the Upbeat & Conventional music dimension, indicating a stronger affinity for familiar, mainstream, and emotionally stable music.

Extending this line of inconsistent results, Schäfer and Mehlhorn reported no consistent or significant correlations between Neuroticism and any music preference dimensions, including Intense & Rebellious music, where an association had been anticipated. Similarly, Rentfrow and Gosling found no substantial correlations between Neuroticism and any of the four music-preference dimensions, further challenging earlier assumptions that neurotic individuals systematically avoid intense or emotionally arousing music. Consistent with these results,

Flannery and Woolhouse observed no differences between individuals high and low in Neuroticism on the acoustic features of music; however, both groups did show preferences for low register and fast tempo.

Collectively, these studies indicate that there is little evidence to support a consistent relationship between Neuroticism and music preferences. Although Langmeyer et al. found a positive correlation for Neurotic individuals towards the Upbeat & Conventional musical dimension, demonstrating that neurotic individuals prefer familiar music, these findings were not supported by other studies.

Big Five Personality Traits	Musical genre (STOMP Model)	Music genre (MUSIC Model)	Acoustic features (tempo, lyrics etc)	Musical characteristics (Valence, Popularity, etc)
Openness to experience	⇑ R&C, I&R, U&C	⇑ Mellow ⇑ Sophisticated ↑ Intense	↑ Dynamics ↑ Piano dynamics ↑ Type of piece	⇓ Valence ⇓ Mainstream ⇓ Danceable
Conscientiousness	(↑/-) U&C (↓/-) I&R	↓ Intense	⇑ Mode ⇑ Type of piece ⇑ Tempo	↑ Tempo
Extraversion	⇑ E&R, U&C ↑ I&R	↑ Contemporary	⇑ Dynamics ⇑ Mode ⇑ Register ⇑ Tempo	↑ Popularity ↑ Danceability
Agreeableness	⇑ U&C	↑ Unpretentious	⇑ Type of piece ↑ Slow tempo music	⇑ Danceability
Neuroticism	(↓/-) I&R (⇑/-) U&C	—	—	—

Legend

⇑	Strong positive correlation
⇓	Strong negative correlation
↑	Moderately positive correlation
↓	Moderately negative correlation
↑	Weak positive correlation

↓	Weak negative correlation
—	No correlation
(⇑/-)	Inconsistent correlations: found a strong positive correlation in some studies whereas found no correlation in the others
(⇓/-)	Inconsistent correlations: found a strong negative correlation in some studies whereas found no correlation in the others
(↑/-)	Inconsistent correlations: found a weak positive correlation in some studies whereas found no correlation in the others
(↓/-)	Inconsistent correlations: found a weak negative correlation in some studies whereas found no correlation in the others

Table 1. Relationship between the Big Five personality traits and music preferences.

Discussion

Personality explains only a small share of variance in music taste, with one clear exception: across foundational studies and a meta-analysis, Openness to Experience is the most reliable correlate. Specifically, listeners who are high on Openness tend to favor music that is novel, complex, and emotionally nuanced. In contrast, Conscientiousness shows at most weak links, such as a small association with Upbeat–Conventional and faster tempo, offering little unique predictive value. Extraversion relates modestly to energetic, socially embedded, beat-driven styles, but meta-analytic effects are small and not uniform, so it should not be treated as a rule. Agreeableness shows small, consistent preferences for cheerful, mainstream, and sometimes danceable music, which is interpreted either descriptively (genre choice) or as an expression of warmth. Neuroticism exhibits minimal, directionally consistent tendencies—namely, a slight pull toward cheerful, familiar styles and away from the most intense music. Overall, Openness matters most; the rest are best viewed as weak tendencies rather than diagnostic markers.

Personality guides musical taste by signaling which kinds of music are likely to fit a person's goals and habits, and that signal can be applied across domains to improve decisions. In recommender systems on music streaming platforms, trait profiles can shape the first few suggestions and offer simple why-this-track explanations. Knowing someone's personality can help to infer their preferences and can therefore contribute to a more accurate recommendation. Additionally, as music plays a role in emotion regulation, and the way people regulate their emotions depends on their personality or their implicit theories of emotion, the music that people use to support their emotion regulation may also be influenced by personality (Ferwerda and Schedl).

Thoma et al. examined the influence of transient personality states on the situational selection of active music. Their work demonstrated that situational music selection follows the principle of emotional congruency: individuals use music as a tool to actively regulate their

emotions in daily life. Depending on the specific emotional context, participants preferred different types of music. This situation-specific selection of emotionally connoted music indicates an attempt by listeners to find individually appropriate music that supports, controls, or changes their current emotions. Highlighting this, Rossi, Oasi, and Colombo found that specific emotion regulation styles significantly impact perceived well-being, as shown through music selection in real-life situations. This insight can be especially valuable in contexts such as music therapy, where patients are encouraged to select the type of music that best supports their emotional states. In fact, music therapy often utilizes patient-preferred music, which is particularly important because it engages and motivates patients; preferred music induces greater brain activation compared to non-preferred music (Brooks Rehabilitation). Taken together, these links serve as evidence-based priors to inform preferences, enabling the personalization of goals based on individuals' needs in real-life contexts.

Future Directions for Research

This paper examined how Big Five traits relate to music preferences broadly. Because personality explains only a small portion of musical preferences, future work should foreground situational context by collecting in-the-moment data on activity, social setting, mood, use of music (such as regulation, focus, or identity), and time of day, and test how these variables mediate or moderate trait–preference links. Within-person and context-controlled designs can reveal when traits matter most and how their effects change across situations. By framing personality as only a small factor in musical preferences, future research can provide more insight into how traits shape listening behavior when considered in real-life contexts. This approach will yield clearer findings that allow for more actionable predictions about music choice.

Additionally, most studies have investigated the effects of the Big Five personality traits on individuals who are high in these traits. While some research has examined the impact on individuals low in these traits, further study is required to enable more meaningful comparisons of music preferences across the full trait spectrum. Future work should focus more closely on individuals lower in traits such as Openness, Conscientiousness, or Extraversion, so that a richer and more balanced picture of personality–music links can be developed.

Of note, this review draws largely on studies of college students, which limits external validity. Future work should recruit a broader adult sample, including young, middle-aged, and older adults across diverse cultural, socioeconomic, and educational backgrounds, preferably via stratified sampling to improve representativeness. This would allow the findings to generalize more confidently to the wider adult population.

Conclusions

This research sought to establish more coherent links between personality traits and music preferences by addressing the methodological inconsistencies that have led to contradictory findings in previous studies. By promoting greater consistency and comparability across research approaches, this review aimed to offer clearer interpretations and justifications

for these discrepancies. Furthermore, it broadened the scope of analysis by incorporating acoustic features such as tempo, rhythm, pitch, and lyrical content, thereby enriching the understanding of music preferences through a more multidimensional perspective.

   The significance of these findings lies in their contribution to understanding the psychological factors that shape individual differences in musical taste. Given that music is a deeply embedded aspect of culture and daily life, examining the relationship between personality and music preferences through diverse methodological frameworks provides a more comprehensive view of the intrinsic influences involved. This research also holds practical implications: it supports more accurate predictions of music preferences and informs applications in areas such as music therapy, personalized recommendation systems, and the functional use of music in clinical settings, where specific genres or acoustic qualities may serve distinct psychological or emotional purposes.

Works Cited

Vella, Elizabeth J., and Gregory Mills. "Personality, Uses of Music, and Music Preference: The Influence of Openness to Experience and Extraversion." Psychology of Music, vol. 45, no. 3, 2017, pp. 338–354. https://doi.org/10.1177/0305735616658957

Rentfrow, Peter J., and Samuel D. Gosling. "The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences." Journal of Personality and Social Psychology, vol. 84, no. 6, 2003, pp. 1236–1256. https://doi.org/10.1037/0022-3514.84.6.1236

Langmeyer, Alexandra, Heinz Guglhör-Rudan, and Christian Tarnai. "What Do Music Preferences Reveal About Personality? A Cross-Cultural Replication Using Self-Ratings and Ratings of Music Samples." Psychology of Music, vol. 40, no. 2, 2012, pp. 164–181. https://doi.org/10.1177/0305735610381597

Moreira Júnior, Luiz Delando Santos, Guilherme Ribeiro Eulálio Cabral, and Giordano Ribeiro Eulálio Cabral. "Correlating Big Five Primary Personality Dimensions with Musical Preferences." Proceedings of the 18th Brazilian Symposium on Computer Music (SBCM), Brazilian Computer Society, 2021, pp. 103–110.

Schäfer, Thomas, and Claudia Mehlhorn. "Can Personality Traits Predict Musical Style Preferences? A Meta-Analysis." Personality and Individual Differences, vol. 116, 2017, pp. 265–273. https://doi.org/10.1016/j.paid.2017.04.061

Flannery, Maya B., and Matthew H. Woolhouse. "Musical Preference: Role of Personality and Music-Related Acoustic Features." Music & Science, vol. 4, 2021, pp. 1–14. https://doi.org/10.1177/20592043211014014

Rentfrow, Peter J., Lewis R. Goldberg, and Daniel J. Levitin. "The Structure of Musical Preferences: A Five-Factor Model." Journal of Personality and Social Psychology, vol. 100, no. 6, 2011, pp. 1139–1157. https://doi.org/10.1037/a0022406.

Rossi, Chiara, Osmano Oasi, and Barbara Colombo. "Personality Characteristics, Music Listening, and Psychological Well-Being: A Systematic and Scoping Review." Research in Psychotherapy: Psychopathology, Process and Outcome, vol. 27, no. 1, 2024, article 742. doi:10.4081/ripppo.2024.742

Setti, Fabio, and Jeffrey H. Kahn. "Evaluating How Facets of Openness to Experience Predict Music Preference." Musicae Scientiae, vol. 28, no. 1, 2023, pp. 143–158. doi:10.1177/10298649231174751.

Cherry, Kendra. "Music and Personality: What's the Connection?" Verywell Mind, 29 Sept. 2022, https://www.verywellmind.com/music-and-personality-2795424.

Cherry, Kendra. "Uses and Gratifications Theory in Media Psychology." Verywell Mind, 10 Apr. 2023, https://www.verywellmind.com/what-is-uses-and-gratifications-theory-in-media-psychology-5217572.

"Empathizing-Systemizing Theory." Goally Neurodiversopedia, Goally, 2023, https://getgoally.com/blog/neurodiversopedia/what-is-empathizing-systemizing-theory/.

McLeod, Saul. "Arousal Theory of Motivation." Simply Psychology, 2018, https://www.simplypsychology.org/arousal-theory-of-motivation.html.

"Register in Music: An Overview." Music Pandit, 2023, https://www.musicpandit.com/resources/articles/register/#:~:text=In%20the%20world%20of%20music,create%20specific%20effects%20in%20music.

"Resource Hub." Brooks Rehabilitation, 2025, https://brooksrehab.org/resource-hub/.

Greenberg, David M., et al. "Just the Way You Are: Music Listening and Personality." Spotify
Research, 7 Dec.
2020, https://research.atspotify.com/2020/12/just-the-way-you-are-music-listening-and-p
ersonality

Cherry, Kendra. "Personality Psychology: The Study of What Makes You …" Verywell Mind, 7
July 2025, https://www.verywellmind.com/personality-psychology-4157179. Verywell
Mind

Cherry, Kendra. "The Big Five Personality Traits." Verywell Mind, 29 Jan.
2025, https://www.verywellmind.com/the-big-five-personality-dimensions-2795422. Ver
ywell Mind

Vinney, Cynthia, PhD. "Uses and Gratifications Theory in Media Psychology." Verywell Mind, 6
Mar.
2024, https://www.verywellmind.com/what-is-uses-and-gratifications-theory-in-media-ps
ychology-5217572.

Nickerson, Charlotte. "Arousal Theory of Motivation: Definition & Examples." Simply
Psychology, 22 Sept.
2023, https://www.simplypsychology.org/arousal-theory-of-motivation.html

Liu, Ziyu. "The Relationship between Music Preferences and Personality Traits." Journal of
Education, Humanities and Social Sciences, vol. 34, July 2024, pp. 6–11.
doi:10.54097/xdecw721.

Young, Craig, and J. Ogn. "Modes and Tonal Preferences in Contemporary Music Studies."
Edited by Daniel Dickman, Journal of Music Theory and Practice, vol. 12, no. 1, 2025,
pp. 45–63.

"Register in Music: An Overview." Music Pandit,
2025, https://www.musicpandit.com/resources/articles/register/#:~:text=In%20the%20wo
rld%20of%20music,create%20 specific%20effects%20in%20music

Ferwerda, Bruce, and Markus Schedl. "Enhancing Music Recommender Systems with
Personality Information and Emotional States: A Proposal." Proceedings of the 2nd
International Workshop on Emotion and Personality in Personalized Systems (EMPIRE
2014), 2014, pp. 1–6.

Thoma, Myriam V., et al. "Emotion Regulation through Listening to Music in Everyday
Situations." Cognition & Emotion, vol. 26, no. 3, 2012, pp.
550–560. doi:10.1080/02699931.2011.595390.

The Intersection of Digital Identity and Financial Technology: Privacy and Surveillance Concerns in Aadhaar-Linked UPI Systems By Kapardhi Pasumarthi

Abstract

This research paper examined perceptions of cybersecurity and privacy risks associated with the use of the Unified Payments Interface (UPI) in India, a widely adopted real-time digital payment system. While UPI has transformed financial transactions through its speed, low cost, and accessibility, concerns regarding data privacy, fraud, and ethical use persist. Using a quantitative research approach, the study employed a structured online survey to collect data from 44 adult UPI users across diverse demographic backgrounds. Descriptive and ex post facto research designs were used to analyze how age, education level, and professional background influence perceptions of fintech-related risks and government regulation.

The findings reveal that age is a significant determinant of concern, with users below 18 and above 50 exhibiting the highest levels of anxiety regarding data disclosure and transactional fraud. Education level shows an inverse relationship with perceived risk: respondents with lower levels of formal education expressed greater concern, while those with higher education demonstrated more confidence and skepticism toward risk severity. Additionally, perceptions of government regulation varied by industry, with homemakers viewing regulation positively, while sales and education professionals were more critical.

Overall, the study highlighted the need for fintech platforms to adopt inclusive design strategies, improve user education, and communicate security measures transparently. Balanced government regulation, combined with user-centric innovation, is essential for building trust and fostering a secure and inclusive digital financial ecosystem in India.

Keywords

Aadhaar-Linked UPI, Fintech, data privacy, surveillance, user concerns, Fintech ethics

Introduction

FinTech applications play a very vital role in all our lives, no matter where we live. FinTech or Financial Technology is a broad and rapidly evolving field at the intersection of finance and technology, two of the important things we cannot live without. It focuses on applying cutting-edge technological innovations to enhance, automate, and streamline financial services and processes (Giglio 601). FinTech covers a vast spectrum of applications that are disrupting and transforming traditional financial services. These applications range from mobile banking to investment and trading, lending platforms to blockchain and cryptocurrencies to personal finance management.

This research paper focuses on a key aspect of FinTech: the Unified Payments Interface (UPI), a real-time payment system developed in India that has transformed the country's digital transaction landscape. UPI enables instant money transfers between bank accounts using a smartphone. It allows users to link multiple bank accounts to a single mobile application,

facilitating seamless fund transfers and merchant payments. UPI's widespread popularity in India can be attributed to its ability to facilitate seamless and instantaneous money transfers, with 24/7 availability. It significantly reduces transaction costs and enhances security through two-factor authentication. These features collectively make UPI a highly convenient and widely adopted FinTech solution for India's large and diverse population. UPI has gained significant traction not only among individual consumers but also within the business sector, where its efficiency, low transaction costs, and ease of integration have made it a preferred payment solution across a wide range of industries. UPI offers businesses several advantages, including instant payment settlements, lower transaction costs compared to traditional methods, and expanded reach to a wider customer base. The security features of UPI also benefit businesses by protecting their transactions and reducing fraud risk. Additionally, UPI simplifies the payment process for merchants, provides valuable transaction data for business insights, and supports the move towards a cashless economy.

Like all transformative technologies, UPI is not without its challenges. Despite the robust security framework established by the National Payments Corporation of India (NPCI), concerns around data privacy and cyber threats persist (Aldboush and Ferdous 6). While the NPCI does not share or sell users' transactional data to private entities, the possibility of malicious actors gaining unauthorized access to sensitive financial information remains a significant risk. Such breaches could potentially lead to the misuse or commercialization of user data, posing serious implications for individual privacy and security.

This research paper explores the ethical and privacy concerns associated with UPI, drawing insights from a survey conducted among 44 individuals from diverse regions and socio-economic backgrounds across India. The findings aim to shed light on public perceptions and apprehensions regarding data security and ethical usage within the UPI ecosystem.

Research Design

A. Research Methodology
This study follows a quantitative research approach because it can accurately reflect the user's concerns on UPI usage through measuring variables, identifying trends, and analysing numerical data. Within this approach, two research designs were chosen: descriptive and ex post facto.

The descriptive design helped capture frequencies, averages, and percentages of demographic information and user perceptions. This created a clear statistical picture of the situation and highlighted key patterns in UPI users' views.

The ex post facto design allowed the study to explore potential cause-and-effect relationships by comparing responses across existing groups, such as different age brackets, income levels, or regions. This way, the study could identify whether certain factors might influence how people perceive cybersecurity risks, without manipulating variables like age or location, which wouldn't be practical or ethical.

B. Data Collection

The primary data collection tool was a structured online survey, chosen for its wide reach and ability to provide quantifiable results. The survey was created in Google Forms and is structured into four parts:

1. Demographic Information: Questions on age, gender, education, income, and region.
2. Usage Patterns: Questions on how often and for what purposes people use UPI.
3. Risk Perceptions: A 5-point Likert scale (strongly agree=5, strongly disagree=1) for participants to express their level of concern about risks like fraud, data privacy, and technical issues.
4. Open-Ended Feedback: An optional question for participants to share additional thoughts or specific experiences.

C. Sample Description

The study focused on adult UPI users in India. The target was to collect 60–100 responses, and the final sample included 44 complete responses. Incomplete responses or answers that were not directly related to the research focus were excluded from the final sample list.

A snowball sampling method was used. The survey was initially shared with known contacts who met the criteria, and they were encouraged to forward it to others. To participate, people needed internet access and an active email or social media account, which naturally excluded those without digital access. Efforts were made to include participants from different age groups, income levels, and regions, although the sample's diversity was limited by the sampling method. Participants had two weeks to respond, with a polite reminder sent after the first week. This approach helped ensure thoughtful and complete responses.

D. Procedure

To maintain consistency and ensure reliable data, the study followed these steps:

1. Finalisation and Testing: The survey was tested with 5 participants to check clarity, functionality, and completion time.
2. Distribution: The live survey link was shared with a brief explanation of the study, a confidentiality note, and a consent statement.
3. Active Collection Period: Responses were collected over 6 weeks, with daily monitoring.
4. Participant Support: The researcher was available via email to clarify any doubts about the survey questions.
5. Follow-up: A reminder was sent midway to improve the response rate.
6. Closure and Data Export: After two weeks, the survey was closed, and the responses were exported into spreadsheets for cleaning and analysis.

E. Limitations

While the study was carefully designed, there are some limitations:

1. Sample Size and Representativeness: With only 44 responses collected through snowball sampling, the findings cannot fully represent all UPI users in India.
2. Geographic Reach: Most participants came from the researcher's extended networks, possibly leading to an overrepresentation of certain cities or states.
3. Digital Bias: People without internet access or digital literacy were excluded, which means the results may not reflect the views of offline users—potentially the group most vulnerable to cybersecurity risks.
4. Demographic Imbalances: Despite efforts to include a variety of participants, some groups (like certain age ranges or income levels) may still be overrepresented.

Being aware of these limitations helps place the study's findings in the right context and highlights areas for future research.

Results

1. Users' Age and Concerns about Fintech Risks

The initial objective was to determine if a user's age influences their concern about various Fintech risks. To ascertain whether older users express greater concern about certain issues compared to younger users, a critical question for UPI adoption and security training was included. This analysis is based on the data presented in Figures 1 and 2.

The novelty of this study lies in examining the different patterns for data disclosure risks across age ranges (Figure 1). The most concern, collectively, was found in the youngest group (Below 18). About 55% of these individuals indicated that they "Agree" or "Strongly Agree" this is a concern. The detailed breakdown revealed that 49% "Strongly Agree", the highest individual figure across all ages appeared for this category. These high levels of aggregated concordance decreased sharply for the middle ages. The 25-35 age range for example, exhibited the most disagreement with a combined percentage of around 50% where users didn't seem to be overly preoccupied about data leaking without consent. Also, the 35-50s are the only age-group in which they did not exhibit a strong opinion in either Strongly Agree or Strong Disagree categories; Neutrals account for more than 40% of them. On the other hand, The 50+ groups were highlighted to be evenly divided among Strong Agree and Agree, with around 40% respectively.
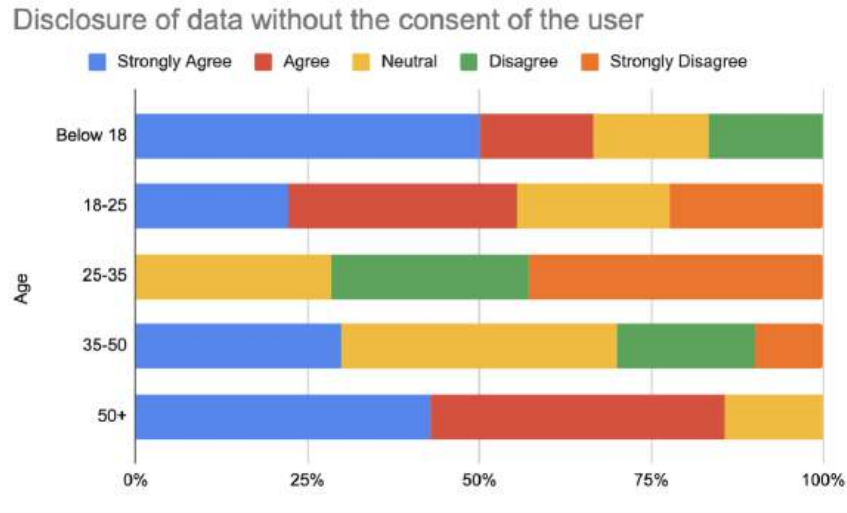
Figure 1: Users' age and their concern about disclosure of data

With regard to the concern related to Transactional fraud/hacking, Figure 2 shows an interesting pattern in which a U-shaped distribution of concern is present. The data showed the results unequivocally, for ages ranging Below 18 and Above 50 where the combined group demonstrated maximum level of concern (Below 18≈70%, Above 50≈65%). This means that the youngest and oldest were most concerned about being hacked or scammed. All other age groups felt least concerned about it in relative terms (compared to their other answers) for the youngest group, 18 to 25 with Neutral as the most frequent response, more than 45% of people selected being neutral. The more moderate levels of agreement were found in the 25-35 and 35-50 age groups, with a combined level of agreement near 45% and 40%. The 35-50 age range, interestingly, had the largest percentage of users actively Disagreeing or Strongly Disagreeing (~20%) that transactional fraud is a concern.

Figure 2: Users' age and their concern about transactional fraud/hacking

2. Users' Education Level and Their Concerns about Fintech Risks

Next, we aimed at querying if a user's Education Level is related to their concerns about various Fintech risks. It would be worthwhile to investigate whether formal schooling affects people's perceptions of the risks associated with UPI-like apps. This section is dedicated to the data presented in Figs 3 and 4.

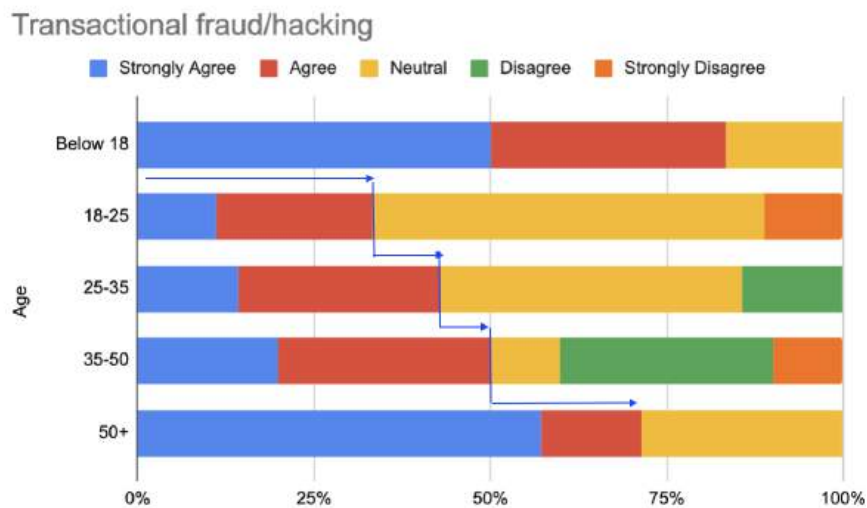Regarding the concern over Disclosure of data without the consent of the user, Figure 3 illustrates a clear inverse relationship with education level and concern. The highest levels of combined concern (Strongly Agree or Agree) are clearly seen in users with No Formal Education and High School Graduates. Both groups show agreement from approximately 70% of respondents. For the Bachelors, Masters, and PhD levels, the combined agreement drops sharply. For example, the Bachelors group has the lowest combined agreement at around 35% , and has the largest segment of Neutral respondents at about 40%. The Masters and PhD groups, while having lower combined agreement than the non-graduates, show a different pattern: the Masters group has a much higher percentage of Strongly Disagree responses (nearly 40%) , while the PhD group shows higher proportions of Agree and Neutral responses compared to the Bachelors and Masters groups.
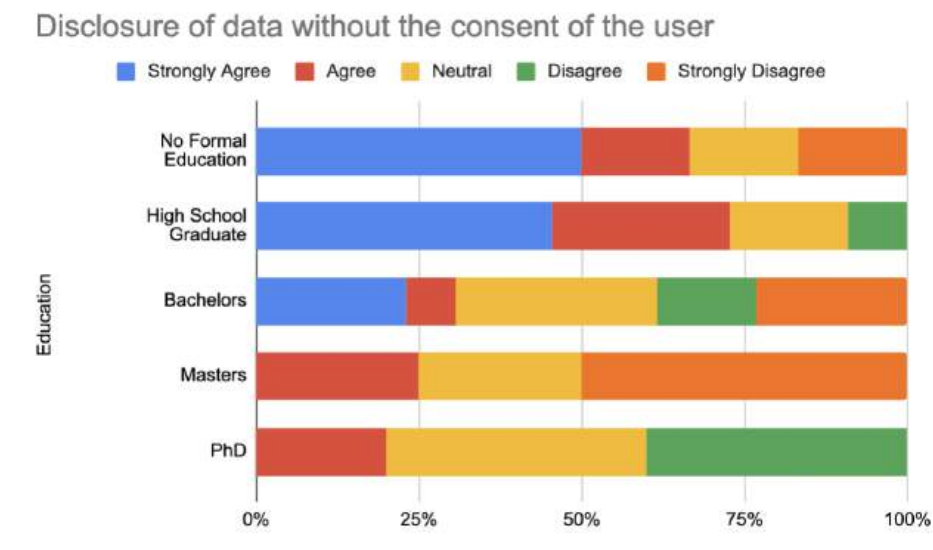
Figure 3: Users' education level and their concern about misuse of data

When looking at the concern over a Data Leak, Figure 4 presents a different but equally compelling pattern based on education level. Users with No Formal Education showed the highest overall combined agreement, with nearly 70% reporting Agree or Strongly Agree. This group, in fact, had virtually no responses in the Disagree or Strongly Disagree categories. In direct contrast, the group with a PhD showed the most pronounced disagreement , with their largest segments being Disagree (about 45%) and Strongly Disagree (about 30%). Furthermore, the PhD group recorded the lowest percentage of Strongly Agree responses at less than 5%. Interestingly, the combined agreement (Agree/Strongly Agree) for High School Graduates and those with a Bachelor's degree fell between 30% and 45%. Specifically, the Bachelors group had

a relatively high proportion of Strongly Disagree responses (about 25%) compared to the High School Graduates (around 10%).
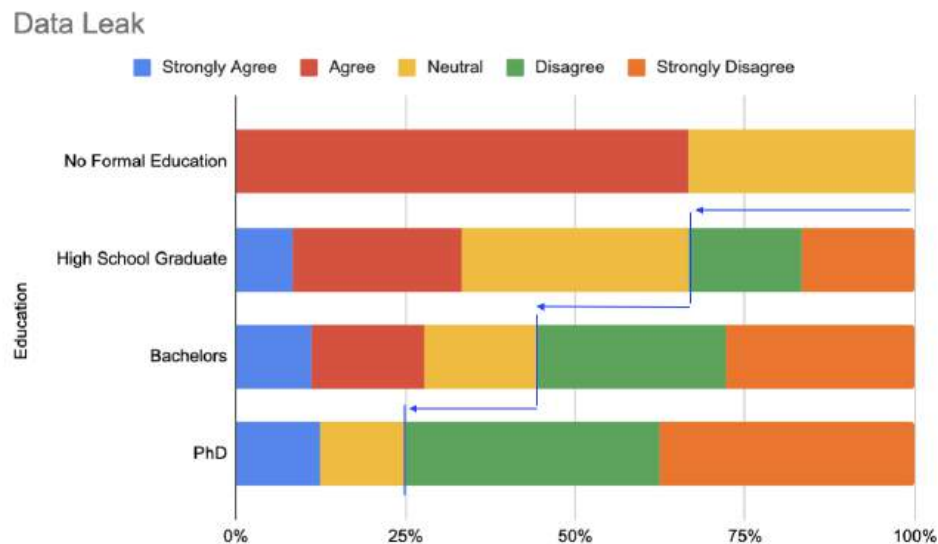
Figure 4: Users' education level and their concern about transactional fraud/hacking

3. Users' Perceptions of Government Regulations of Fintech Apps by Industries

 Our final analytical goal was to determine how a user's Industry of employment relates to their Perceptions of Government Regulations of Fintech Apps. We used a 5-point scale (where 5 is "strongly agree regulation is beneficial," and 1 is "strongly disagree") to capture these views. Figure 5 provides the mean and median scores for each industry's perception of government regulation. These averages provide a snapshot of the perceived benefits and risks of government policies across business sectors. We found a clear distinction between highly supportive and highly skeptical groups. Participants working as a Home Maker had the highest average score, with both a mean and median of 4.0. The industries showing the least favorable, or most skeptical, perception of government regulation were Sales and Education. The Sales industry reported the lowest scores, with a mean of 2.0 and a corresponding median of 2.0. The Education industry was close behind, with a mean of 2.5 and a median of 2.5. All other industries fell in the middle range, suggesting a more neutral or slightly favorable view, with their scores clustering around the midpoint. For instance, Manufacturing & Construction had the highest mean among the middle group at approximately 3.43, while Services had the lowest at about 2.67, but all three—including Agriculture & Natural Resources—shared an identical median score of 3.0. The Technology & Transportation sector was slightly higher, reporting a mean of 3.25 and a median of 3.5.

Industry	mean	median
Agriculture & Natural Resources	2.8	3
Manufacturing & Construction	3.428571429	3
Services	2.666666667	3
Technology & Transportation	3.25	3.5
Home Maker	4	4
Education	2.5	2.5
Sales	2	2

Figure 5: Users' Industry and their perceptions of government regulation of fintech apps

Discussion

The findings of this research showed that people's concerns about sharing personal data in fintech apps are strongly influenced by their age. Users under 18 and those over 50 tend to feel much more worried about privacy compared to middle-aged users. This is likely because younger users are still developing financial independence, while older users may have limited exposure to modern financial technologies. As a result, the risks of using digital platforms can feel much bigger than the benefits for these groups.

To reduce this fear, fintech companies should focus on helping users feel more comfortable and informed. Simple steps such as interactive tutorials, short videos, guided onboarding, and built-in help features can make a significant difference. When users understand how a platform works and how their data is protected, their confidence naturally increases.

Education level also plays an important role in shaping how secure users feel. People with higher levels of education tend to be less worried about data privacy because they are often better equipped to understand digital systems and manage security settings on their own. On the other hand, users with little or no formal education may find these platforms confusing or intimidating. For them, fintech companies need to simplify instructions, offer clear explanations, provide easy access to customer support, and send reassuring updates that help build trust over time.

Professional background further affects how users view government involvement in fintech. Homemakers often see government regulation as a form of protection that makes digital finance safer. In contrast, sales professionals may view regulation as restrictive, fearing it could limit flexibility or slow down business. Educators generally take a balanced approach, recognizing both the benefits and drawbacks of regulation. Because of these differing perspectives, fintech companies must carefully balance compliance with innovation and communicate their policies in ways that resonate with different professional groups.

Conclusion

Overall, this study makes it clear that trust and privacy concerns in fintech are not the same for everyone. Instead, they are shaped by a combination of age, education, and professional background. To meet these varied needs, fintech platforms must focus on user-friendly designs that prioritize clarity, education, and transparency. Making technology easier to understand is just as important as making it secure.

At the same time, cooperation between governments and fintech companies is essential. Strong regulations are necessary to protect users, but they should not come at the cost of accessibility or innovation. When done right, regulation can support trust without limiting the freedom that makes digital financial services so useful.

Addressing these differences creates a more inclusive financial ecosystem—one where users of all ages and backgrounds feel empowered to manage their money safely and independently. Currently, Fintech is a developing and evolving subject with lots of untouched depth that is yet to be studied and implemented by many fintech companies. More research on the actual usage data of Fintech apps would help to establish a greater degree of accuracy on the matters of privacy and ethical concerns.

Works Cited

Abed, Ali Kamil, and Angesh Anupam. "Review of Security Issues in Internet of Things and
   Artificial Intelligence‑driven Solutions." SECURITY AND PRIVACY, vol. 6, no. 3, May
   2023, p. e285. DOI.org (Crossref), https://doi.org/10.1002/spy2.285.

Adedoyin Tolulope Oyewole, et al. "DATA PRIVACY LAWS AND THEIR IMPACT ON
   FINANCIAL TECHNOLOGY COMPANIES: A REVIEW." Computer Science & IT
   Research Journal, vol. 5, no. 3, Mar. 2024, pp. 628–50. DOI.org (Crossref),
   https://doi.org/10.51594/csitrj.v5i3.911.

Aldboush, Hassan H. H., and Marah Ferdous. "Building Trust in Fintech: An Analysis of Ethical
   and Privacy Considerations in the Intersection of Big Data, AI, and Customer Trust."
   International Journal of Financial Studies, vol. 11, no. 3, July 2023, p. 90. DOI.org
   (Crossref), https://doi.org/10.3390/ijfs11030090.

Giglio, Ferdinando. "Fintech: A Literature Review." EUROPEAN RESEARCH STUDIES
   JOURNAL, XXIV, no. Issue 2B, June 2021, pp. 600–27. DOI.org (Crossref),
   https://doi.org/10.35808/ersj/2254.

Liu, Qianhua, et al. "Fintech Research: Systematic Mapping, Classification, and Future
   Directions." Financial Innovation, vol. 10, no. 1, Jan. 2024, p. 24. DOI.org (Crossref),
   https://doi.org/10.1186/s40854-023-00524-z.

Rizinski, Maryan, et al. "Ethically Responsible Machine Learning in Fintech." IEEE Access, vol.
   10, 2022, pp. 97531–54. IEEE Xplore, https://doi.org/10.1109/ACCESS.2022.3202889.

The Price of Discrimination: Legal and Reputational Risks of Discriminatory Practices
By Armaan Vir Kumar

Abstract
Employment discrimination is one of the most common issues that affect firms today. Employment discrimination is prevalent even in the presence of laws such as Title VII of the Civil Rights Act protecting against discrimination based on race, color, religion, sex, or national origin, the Age Discrimination in Employment Act, and the Americans with Disabilities Act. The purpose of this essay is to analyze the impact of breaching these anti-discrimination laws and the exposure of companies to discrimination-related civil lawsuits, and the impact of such lawsuits on the businesses' reputation. Such impact is detrimental to the company's sustainability as businesses lose credibility and reputation. The study largely contributes to the existing body of knowledge whose most works seem to be concerned with the legal issues only. The present essay draws from major lawsuits filed against Abercrombie & Fitch, Google, Walmart, and Denny's to show that incurring such lawsuits has legal and reputational implications. The research shows that the companies have lost a significant amount of money owing to lawsuits against them, and have publicized their sponsored discrimination based on race, religion, age, disability, or pregnancy, and have lost their reputation. The paper also shows that these companies, after facing discrimination lawsuits, have suffered a lot and have been able to heal the wounds by adopting transparency, inclusiveness, and sustainability.

Keywords
Employment Discrimination, Anti discrimination laws, workplace inclusion, legal compliance, corporate accountability

1. Introduction
Worldwide, the hiring and onboarding of employees is a fundamental part of running a business (WECGlobal). On the surface, it appears to be a fairly simple process consisting of submitting a resume and interviewing for a position. However, there is far more to it than that. There is a complex network of rules and legal responsibilities that govern the process, including anti discrimination laws. While there must be regulations surrounding the fairness and equal opportunity for discrimination, one must first ensure that the individuals in a given pool of candidates are qualified for the position.

The aim of this paper is to discuss some of the more prominent discrimination laws in the current landscape, especially pertaining to the United States. The following sections examine not only the law itself, but also the ramifications of failing to comply with the law on the part of the employer. Moreover, the paper will analyze the mechanisms and steps that businesses can utilize to improve or restore their position post-offense.

The paper discusses the legal implications of anti discrimination laws in the United States in regard to the discriminatory practices in hiring that result in the affected business being

sanctioned. It also considers the ways in which the affected businesses can prevent, or mitigate, the self-inflicted negative consequences through what can be termed as genuine positive changes.The following sections present a structured analysis of key legislative instruments, their implications for the employers, and the broader impact of violations upon business reputation and its operational viability.

2. Key Legislative Frameworks Governing Employment Discrimination Legal Repercussions for Non-Compliance

2.1 Title VII of the Civil Rights Act of 1964

Enacted as a cornerstone of American civil rights legislation, Title VII of the Civil Rights Act of 1964 prohibits employment discrimination based on race, colour, religion, sex (inclusive of gender identity, sexual orientation, and pregnancy status which were added to the title slightly later), and national origin. This statute promotes diversity and inclusivity in the workplace and applies to a broad spectrum of employers.

However, there are notable exclusions. The statute does not extend to independent contractors, American citizens employed abroad by non-U.S.-controlled firms, or foreign nationals employed overseas by U.S.-based enterprises.

The anti discriminatory laws were mainly introduced to provide the citizens with equal rights and opportunities following the civil war that had taken place in the country, where thousands had been enslaved and inequality was rampant. It also followed the principle of American democracy which aims to ensure pursuit of equality and the protection of one's individual rights.

2.2 The Age Discrimination in Employment Act of 1967 (ADEA)

The ADEA was established to protect individuals aged forty and above from discrimination in hiring, discharge, compensation, or terms and conditions of employment. It applies to private employers with twenty or more employees, as well as governmental and labour organizations (US EEOC). (n.d.-a)

Importantly, an employee alleging discrimination under the ADEA must demonstrate that age was the primary motivating factor behind the adverse employment action. In certain circumstances, employers may offer severance agreements that include waivers of ADEA claims, provided such waivers are made knowingly and voluntarily.

2.3 The Americans with Disabilities Act of 1990 (ADA)

The ADA is a landmark piece of legislation that affords civil rights protections to individuals with physical and mental disabilities. It obligates employers with fifteen or more employees to provide reasonable accommodations and prohibits discrimination in all employment practices. Notably, the ADA Amendments Act of 2008 clarified key provisions. For

instance, individuals whose impairments are mitigated by medication or corrective lenses are not automatically deemed disabled under the Act.

2.4 Violations of Title VII

According to Thomson Reuters (2024), employers found in breach of Title VII may face a range of legal penalties, including but not limited to back pay, front pay, compensatory and punitive damages, and injunctive relief. Under 42 U.S.C. 1981a, claimants must demonstrate that discrimination was intentional in order to be eligible for such remedies. The extent of compensatory and punitive damages is capped based on the size of the employer. Courts are empowered to issue injunctions requiring the cessation of discriminatory practices and the implementation of appropriate remedial actions.

2.5 Violations of The Age Discrimination in Employment Act

The US EEOC (n.d.-b states the legal consequences for breaching the ADEA, which include reinstatement, front pay, back pay, and liquidated damages, which are often equal to back pay in cases of willful violations. A business found guilty of age discrimination may also come under scrutiny by the Equal Employment Opportunity Commission (EEOC), which can initiate investigations into broader employment practices.

2.6 Violations of the The Americans with Disabilities Act

The U.S. Equal Employment Opportunity Commission also highlights the repercussions of the violations of The Americans with Disabilities Act of 1990. According to them, the individuals subjected to discrimination under the ADA must typically file a charge with the EEOC within 180 days, though this period may be extended to 300 days in certain states. Remedies include hiring, reinstatement, back pay, compensatory damages, and reasonable accommodations. Prevailing complainants may also recover attorney's fees.

3. Broader Implications for Businesses

3.1 Legal and Financial Repercussions

Discriminatory practices not only expose organizations to direct legal penalties but also compel them to bear substantial legal fees and settlement costs. While the form of punishment varies by jurisdiction and case specifics, financial remuneration to the aggrieved party remains the most common outcome.

3.2 Damage to Reputation and Public Trust

The damages that companies face for losing reputation and being accused of discriminatory practices can cause harm in more ways than one. Losing trust from the general public, i.e. the public outcry, can lead to withdrawal and public boycott of their services, loss of

partnerships and decreasing trust and confidence from their investors, resulting in more harm than just monetary damages.

One can argue the Worcester Art Museum case in Massachusetts as stated by Hartig in 2021. A Muslim security guard was separated from his position and the circumstances involving the case were rather weak. After the 9/11 incident, the EEOC investigated the case and the museum was forced to pay the 60,000 settlement amount as a result of discriminatory practices. This scenario not only explains the loss of reputation because of discriminatory practices, but it also explains the negative economic and social impact that can result from a culturally ignorant practice of discrimination in the workforce.

4. Historical Case Studies and Real-World Business Examples

4.1 Abercrombie & Fitch: Discrimination Based on Religion

In 2015, the United States Supreme Court ruled in EEOC v. Abercrombie & Fitch Stores, Inc., a landmark case concerning religious discrimination in hiring practices. The case involved a Muslim teenager who wore a headscarf (hijab) during a job interview. Although she was otherwise qualified, Abercrombie declined to hire her, citing its "Look Policy," which prohibited headwear.

The Court found that employers cannot take employment applicants' religious practices into account when making hiring and employment decisions. The Court also noted that in order to be liable, one does not even need to have actual knowledge of the religious requirement; mere suspicion is enough. This caused businesses to have to adjust their policies in such a way that, although neutral on the surface, impacted protected religious groups more significantly.

4.2 Google: Age Discrimination Allegations

Google has been sued several times because of age discrimination. The company settled in 2019 for $11 million to over 200 class action applicants over age discrimination, which includes job applicants over the age of 40, whom the company allegedly would not hire. Although Google settled to avoid further escalating the issue, the case was a reflection of ongoing age discrimination in the tech industry.

More Information on data transparency also applies to Thiscase. Increasingly, companies are expected to analyze their hiring data and results to evaluate if their policies are age neutral. Google, and other companies, increased their advertising and recruiting to promote age diversity due to the need to positively impact their reputation.

4.3 Walmart: Disability and Pregnancy Discrimination

Walmart has faced several discrimination lawsuits, particularly with respect to disability and pregnancy. The EEOC filed a national hiring discrimination lawsuit, which led to a $20 million settlement with Walmart in 2020.The aforementioned case alleged that Walmart failed to

accommodate applicants who were deaf by not making arrangements in terms of sign language interpreters during interviews.

A separate case involved the company being sued for not offering lighter duty assignments to pregnant employees. In 2014, Walmart changed its policy to offer reasonable action. These lawsuits, however, pointed out the fallout of not following the Americans with Disabilities Act and the Pregnancy Discrimination Act.

All these incidents show the legal and public relations fallout from failing to accommodate employees as required by law. They also show the expectations of companies to take on these legal responsibilities and, in addition, to show some degree of social responsibility.

4.4 Denny's Restaurants: Racial Discrimination

Denny's Restaurants in the 1990s was, arguably, the most well-known for its racially discriminatory practices, such as not serving African Americans, and if they did, taking long periods of time to do so. This culminated in a class action lawsuit, and in 1994, Denny's settled for $54 million—the biggest settlement of such a class action lawsuit at the time.

This case stands as a historical marker in the evolution of civil rights enforcement in the private sector. Denny's underwent vast structural and operational changes in the organization, not limited to, but including, different diversity training, programs aimed at recruitment of minorities, and the appointment of African Americans in senior executive positions.

This move was very expensive for Denny's, but in the long run, public perception had a positive turn, and proved to others businesses that recovering from public backlash is possible if they engage in sustained, genuine efforts towards reform.

4.5 The Pattern

All the companies were sued for employment discrimination in one form or other for considering factors like religion, age, disability and pregnancy irrespective of the size and magnanimity of the organization. All the companies involved sustained a great loss regarding their public image, which in consequence, may lead to other loss magnitudes. In addition, a number of companies were even penalized for the lawsuits which were alleged against them. Image wise, and for the sake of society, better changes were brought, although it may be expensive like the case of Denny's. These discrimination lawsuits illustrate how companies lost in employment process discrimination.

5. Lessons Learned and Policy Recommendations

The previous case studies, when viewed together with the earlier described regulatory framework, provides a number of valuable insights for public and private entities. In a world that is increasingly interwoven and characterized by social consciousness, the views of employers have expanded to go beyond the regulatory framework. There seems to be a social responsibility regarding employment, which should be fair, just and inclusive. In this context, the following

insights and recommendations concerning policies are being put forward for public, private employers, and for human resource and labor relations practitioners.

5.1 Prevention is Preferable to Penalty

One of the most evident conclusions to be drawn is that proactive prevention of discriminatory practices is invariably more effective, both in economic and reputational terms, than reactive redress following a violation. The resources to implement preventive measures are within range compared to the consequences of the litigation costs, punitive damages, governmental questioning, and public questioning of the entity.

Companies are encouraged to provide company-wide anti-discrimination training to employees of all levels, including front-line employees, employees who are decision-makers for hiring and firing, and those who perform supervisory and disciplinary functions. Such training should not be a one-time event but should be ongoing and customized to incorporate change in legal requirements and complexities of the situations.

In addition, it is necessitated that there be a regular and systematic internal review of the data pertaining to hiring, data on employee promotions, and data on the results of employee surveys which measure the internal culture of the organization. Such audits should be designed to unsurface patterns of unrecognized imbalances or inequitable structures that would more likely come to the attention of the organization if litigation were to occur. Such audits might be in even more comprehensive and thorough processes which go beyond data gathering, legal review, and design of organizational policies, but would also aim to provide employees with various ways to file grievances which would be appropriately protected from retaliation.

5.2 Diversity Must Be Meaningful, Not Cosmetic

Inclusion and equity should not be a superficial exercise. Courts, regulators, and the public are able to distinguish between real equity and equity that is a charade, and that inequitable practice can no longer be allowed to continue. Simply, a set of quotas for a diverse employee base can no longer exist while the culture and policies of the organization are exclusionary and isolatory of a particular group.

Thus, firms need to foster a culture of real inclusion, where everyone, of all backgrounds, is welcomed, and where people feel able to add their voice and are protected and valued. Areas that need to be factored in are: affinity group(s), mentoring, leadership development, and the use of respectful language. There must also be, and just as importantly, a focus on the composition of the most senior roles; inclusivity is not simply about having a diverse intake to a job advocacy (especially advocacy position) but also to that job's tier(s) with the major decision-making authority, and to leadership.

Additionally, employers must ensure that they routinely amend job descriptions, and performance and promotion criteria, in order to not create unintentional biases that disadvantage particular groups. Any system in place that has the net effect of excluding and/or disadvantaging

people on the basis of their protected characteristics, even where a primary purpose is not to discriminate, must be changed.

5.3 Transparency and Accountability Matter

Third on this list is the need for clarity and accountability when making employment decisions. Many discrimination lawsuits are not the result of discrimination alone but are the result of an absence of an explanation of a credible reason and procedural fairness that is verifiable by the employee.

In this light, there is a need for employers to apply consistent hiring and assessment processes that are based on measurable criteria rather than subjective criteria. Assessment forms, interview score sheets, and reasons for hire should be collected and kept on file for review in the event of an audit or legal challenge.

In addition, there should be accommodation processes that are accessible to all employees and applicants that have been documented and communicated. Where there is a refusal of an accommodation, there should be a written explanation that is legitimately based on business necessity.

To have an ethically responsible corporate culture, accountability can be enhanced through the use of mechanisms for anonymous reporting or by implementing a system of third-party oversight. Where possible, community trust can be strengthened by the organization through publicly accessible audit reports and independent accountability on diversity within the organization.

5.4 Long-Term Investment in Ethical Practices

Finally, the adoption of non-discriminatory and inclusive employment practices should not be viewed as a legal obligation alone, but as a long-term investment in organizational health and sustainability. Numerous studies have demonstrated that diverse and inclusive workplaces enjoy greater innovation, employee satisfaction, and public goodwill.

In this light, corporate leadership must integrate equality into the very mission and vision statements of their organizations. Boards of directors and senior management must treat anti-discrimination not as a peripheral legal concern but as a core business strategy.

The legal, ethical, and financial imperatives all point in the same direction: equality, when truly embedded, is not only just but profitable.

6. Limitations

1. Scope: This study only pertains to those cases which are applicable under the US anti discriminatory laws. It fails to cover the problem in countries outside the United States.

2. Selection of Studies: The studies only cover large companies such as Dennys. Studying the same problem in small sector enterprises will widen the scope of the study.

3. Focus on Business and Legal Perspective: The study focuses on the corporate and legal perspective of the problem at hand. By studying the perspective of the employees, it could help understand the problem in greater detail.

4. Focus on Qualitative Data: The study is primarily qualitative, as there is little quantitative data. Including more quantitative data would aid in drawing better conclusions from the study.

7. Conclusion

The present study has brought to light the key legislative tools available to combat employment discrimination, as well as the penalties arising from the employment discrimination legislation. Employment discrimination, on the basis of race, age, and disability, is unlawful as far as the federal legislation in the USA is concerned. The primary employment discrimination legislation within the USA seeks to promote and protect equal employment opportunities and employment practices.

Legal sanctions—most often financial in nature—are the standard means of redress for victims of discrimination. However, as this paper has argued, the reputational harm sustained by violators may prove far more injurious in the long term. Companies must therefore strive not merely to comply with the letter of the law but to embrace its spirit, cultivating workplace cultures rooted in fairness, dignity, and inclusivity.

Works Cited

"Title VII of the Civil Rights Act: The Basics You Should Know." Thomson Reuters, 6 Aug.
    2024, legal.thomsonreuters.com/en/insights/articles/what-is-title-vii-civil-rights-act.
    Accessed 15 Mar. 2025.

Barnes, Patricia. "Deja Vu: Google Settles Age Discrimination Lawsuit for $11 Million." Forbes,
    20 July 2019,
    www.forbes.com/sites/patriciagbarnes/2019/07/20/deja-vu-google-settles-age-discriminat
    ion-lawsuit-for-11-million/.

Hartig, Hannah. "Two Decades Later, the Enduring Legacy of 9/11." Pew Research Center, 2
    Sept. 2021,
    www.pewresearch.org/politics/2021/09/02/two-decades-later-the-enduring-legacy-of-9-1
    1/.

U.S. Equal Employment Opportunity Commission. "Age Discrimination in Employment Act of
    1967." EEOC, www.eeoc.gov/statutes/age-discrimination-employment-act-1967.
    Accessed 15 Mar. 2025.

U.S. Equal Employment Opportunity Commission. "Fact Sheet: Age Discrimination." EEOC,
    www.eeoc.gov/laws/guidance/fact-sheet-age-discrimination. Accessed 15 Mar. 2025.

U.S. Equal Employment Opportunity Commission. "The ADA: Your Employment Rights as an
    Individual With Disability." EEOC,
    www.eeoc.gov/publications/ada-your-employment-rights-individual-disability. Accessed
    15 Mar. 2025.

World Employment Confederation. World Employment Confederation,
    wecglobal.org/world-employment-confederation-global. Accessed 15 Mar. 2025.

"Influence of Flame-Retardant Additives on Polymer Combustion and Mechanical Properties" By Matthew Tam

Introduction

Polymers are widely used materials in almost every aspect of life. From the circuit boards in computers to the window seals keeping houses warm, their use is ubiquitous and eclectic (Hsissou et al.). However, these polymers have a strong underlying risk: they are extremely flammable.

Commonly used polymers are hydrocarbons, so are inherently very flammable and release heat when burned. For example, polyethylene (PE) and polypropylene (PP) have respective heats of combustion of about 43.1 MJ/kg - 44.6 MJ/kg, which are comparable to gasoline's heat of combustion of about 43.7 MJ/kg (Barbrauskas et al.). The main driving factor of the heat of combustion of hydrocarbon polymers is the percentage of oxygen and other atoms, such as nitrogen, in the molecule (Barbrauskas et al.). The lower a polymer's oxygen content, the greater its heat of combustion. The opposite is also true: the greater the content of non-carbon and non-hydrogen atoms in a polymer, the lower its heat of combustion. PE and PP both lack oxygen in their respective chemical formulas, which contributes to their high heats of combustion.

Inherently flame-retardant polymers, such as polyphenylene sulfides, do exist; however, they are often much more expensive than other commercially available polymers. Figure 1 below shows the price per pound versus the heat release capacity of many commercially available polymers. Figure 1 also shows a clear correlation between the polymer's price and its flame retardancy. Most commodity polymers have a high heat release capacity and therefore sustain combustion when ignited. Whereas very few engineering or specialty polymers sustain combustion when ignited. Using inherently flame-retardant polymers would be ideal; however, it is impractical, as it would drastically increase the cost of goods. As shown in Figure 1 below, self-extinguishing polymers are frequently multiple times more expensive than cheaper, less flame-retardant polymers. As shown by Figure 1, most commodity polymers are in the 0.1-to-2-dollar price per pound range, whereas engineering polymers are mostly in the 2-to-50-dollar price per pound range. Additionally, cheaper flame-retardant polymers may not exhibit as desirable properties as cheaper non-flame-retardant polymers.
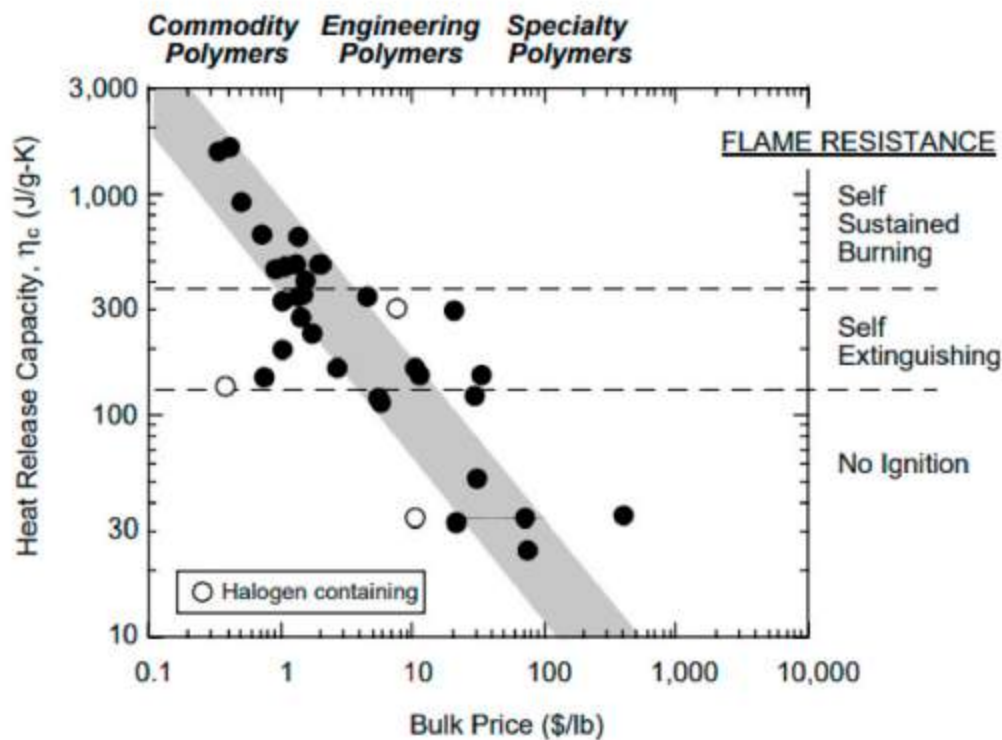
Fig. 1. Heat release capacity, cost and combustion behaviour of commercial polymers [1].

Figure 1. Graph of polymer cost, heat release capacity, and flame resistance (Lyon et al.)

Fire disasters have caused significant damage to human life and property. In 2020, fire damages resulted in an estimated total of 4 billion yuan in China and about 1183 deaths (Ministry of Emergency Management). Polymers, being very flammable, contributed to fire disasters by posing a significant fire hazard. When polymers burn, they produce large amounts of smoke and heat, which not only feed the fire but also suffocate people in the burning building.

To lessen the risk of fire and the amount of fire disasters that occur, additives have been researched and developed to make polymers flame retardant while keeping costs low. Properties of desirable flame-retardant additives include high flame-retardant efficiency, low smoke production, and non-toxicity, among others. Ideally, the additive would enhance the polymer's mechanical properties, or at least not inhibit them. However, as discussed later in this paper, this is usually not the case, and flame-retardant additives typically degrade the mechanical properties of a polymer (Liu et al.). This paper analyzes how different flame-retardant additives, such as inorganic, organic, and nano-fillers, affect the flammability and mechanical properties of polymers.

Combustion

The combustion of polymers is a complex process that involves four main stages: ignition, pyrolysis, combustion, and feed-back. Figure 2 below shows a basic diagram of the four stages in the combustion process. The heat from ignition causes the polymer to pyrolyze (decompose) into flammable gases and vapors. This pyrolysis step is endothermic, which is why the polymer does not spontaneously combust at room temperature. Then, combustible gases and vapors undergo exothermic oxidation reactions, producing •OH and •H radicals that help sustain the flame. The large amount of heat produced starts the cycle again by pyrolyzing the polymer and producing more flammable gases.
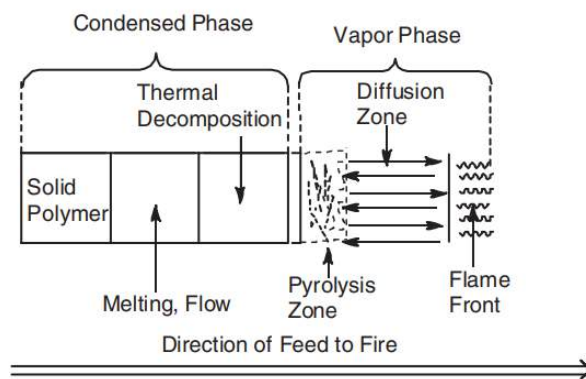

Figure 2. Diagram of the process of polymer combustion. (Morgan et al.)

During combustion, there are three main phases in which the combustion of polymers occurs: the condensed phase, the mesophase and the vapor phase. During the condensed phase, the polymer is in the solid state and undergoes pyrolysis, producing volatile gases. During the vapor phase, the volatile gases react with oxygen in violent exothermic reactions and create •OH and •H radicals to continue combustion. The mesophase is a less-discussed phase of combustion that occurs between the condensed and vapor phases. During the mesophase, the polymer exists in a semi-fluid state, allowing heat and mass transfer between the environment and the polymer (Shen et al.).

Measurements

There are many tests designed to measure the flammability of materials, such as the limiting oxygen index (LOI) test, the UL-94 Test, and others. Each test has limitations, which will be covered later in this section. Nevertheless, these tests are useful in determining the overall flame retardancy of polymers and their additives. Often, a combination of these flammability tests is used to make stronger conclusions on the flame retardancy of the studied additive.

The LOI test is defined as the minimum oxygen concentration required to maintain a top-down combustion for 3 minutes or until 5 cm of the material is consumed (Liu et al.; Shen et al.). The value given by the LOI test is the percent oxygen concentration; the higher the LOI value, the greater the material's resistance to combustion. However, this test has limitations, as it leaves out variables such as heat release rate and time to ignition (Shen et al.). Materials that

have an LOI value greater than 28% are generally regarded as "self-extinguishing," and the additives that help achieve this value are considered effective flame retardants (Araby et al.).

Underwriters Laboratories developed the UL-94 test to test the flammability of plastics. This test is one of the most common methods for determining the flammability of plastics and the spreadability of the flame (Shen et al.). A polymer can achieve one of six classification levels: HB, V-2, V-1, V-0, 5VB, and 5VA, from lowest flame retardancy to highest. The qualifications for these classifications can be found on the UL website (UL Solutions). During the test, a 12.7 cm by 1.27 cm sample is burned over a 20 mm 50 W blue flame. This test measures the time it takes for the polymer to extinguish after 10 seconds of flame exposure (T_1) and the time it takes for the polymer to extinguish after a subsequent 10 seconds of flame exposure (T_2). Additionally, the UL-94 test measures the dripping behavior of the polymer when combusted. For a polymer to achieve a V-0 rating, for example, it must burn for less than 10 seconds and produce no drips (UL Solutions; Marti et al). One possible limitation of this test is that it gives a categorization rather than a quantitative metric.

Discussion

Inorganic Flame Retardants

Metal Hydroxides

One of the main categories of the inorganic additive flame retardant, which have no carbon in their chemical backbone structure, is the metal hydroxide group (Liu et al.; Shen et al.). Metal hydroxide compounds, as the name suggests, contain a metal bonded to a number of hydroxide ions. Two common metal hydroxide compounds used are aluminum trihydroxide ($Al(OH)_3$) and magnesium dihydroxide ($Mg(OH)_2$).

Metal hydroxide flame retardants work by endothermically decomposing when exposed to heat. The metal compounds decompose into a metal oxide and water, which lowers the temperature of the polymer and generates water vapor and carbon dioxide, reducing oxygen concentration. The metal oxide layer, further, produces a film on top of the polymer that helps flame retard the polymer. As shown in figure 3 below, the char layer helps insulate the polymer (in the thermal degradation zone) from the combustible gases, oxygen and heat.
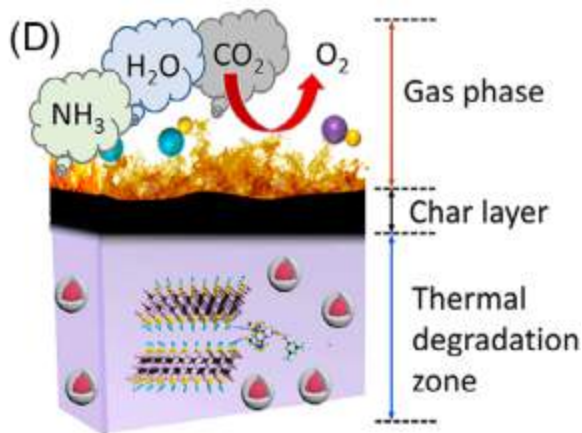
Figure 3. Diagram of char layer mechanism. (Wang et al.; Liu et al.)

These metal-hydroxide compounds are non-toxic, environmentally friendly, cheap, and release little smoke (Liu et al.; Shen et al.; CRC Press; Royal Society of Chemistry). Despite their numerous advantages, they are very inefficient, in that high concentrations of metal compounds are required to sufficiently flame retard a polymer (Liu et al.; Shen et al.). Usually, these compounds are added at 40-60 wt%. Additionally, this high loading significantly reduces the mechanical strength of the polymer(Liu et al.). For example, when magnesium hydroxide was added at 50 wt%, it decreased tensile strength by ~62% (Araby et al; Camino et al).

To cope with this problem and to improve the dispersibility of metal hydroxide compounds, Lee et al. added silane coupling agents to magnesium hydroxide (Lee et al.). This resulted in better dispersion of magnesium hydroxide in the polymer and increased tensile strength and elongation when the coupling agent was added. Figure 4 shows SEM pictures of a before and after tensile test of polypropylene-based polymers mixed with magnesium hydroxide and varying levels of silane coupling agents.
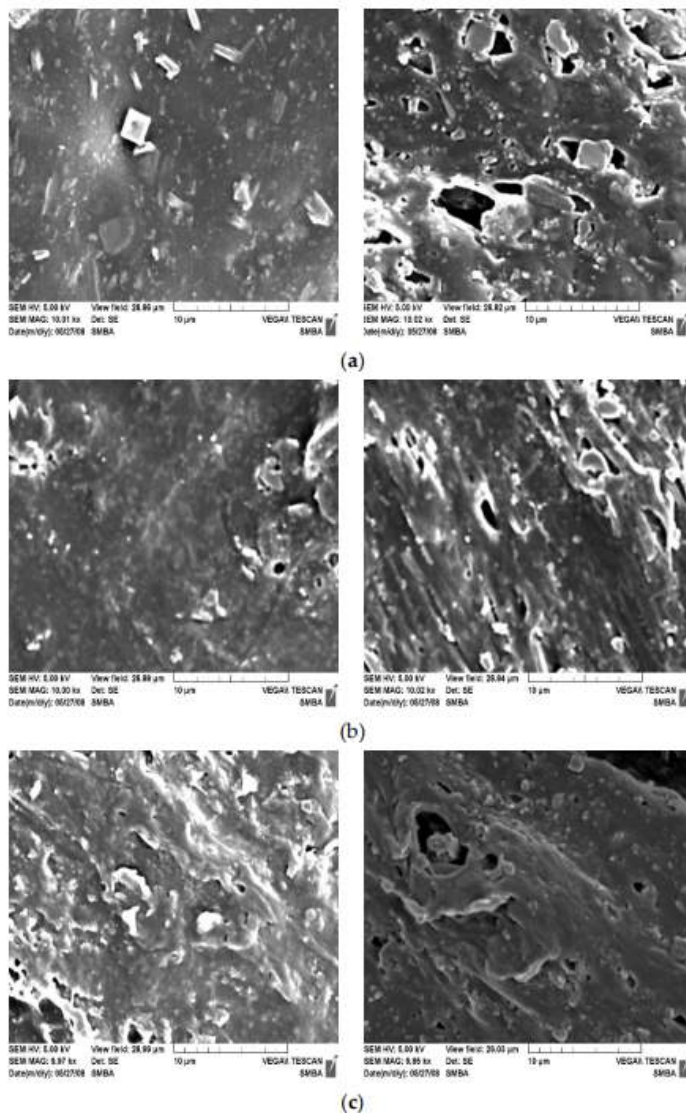
Figure 4. SEM images before and after a tensile test of a sample's surface. (a) Mg(OH)$_2$ without treatment; (b) Mg(OH)$_2$ with silane treatment; (c) Mg(OH)$_2$ with excessive silane treatment (Lee et al.)

Phosphorus Flame Retardants

Another common inorganic flame retardant used is the phosphorus-based flame retardant. Widely used phosphorus flame retardants include ammonium polyphosphate (H(NH$_4$PO$_3$)$_n$OH) and red phosphorus. Phosphorus additives are known for being very good flame retardants and are synergistic with other additives.

Red phosphorus is known for being very efficient, non-toxic, and having synergy with other flame retardants (Shen et al.). However, it absorbs water from the air and creates phosphoric acid and highly toxic PH$_3$ gas when exposed to air for extended periods of time. Additionally, it is not very compatible with resin and is difficult to disperse. When red phosphorus is added to resin, the viscosity of the resin increases, which makes processing more

difficult and inhibits the mechanical properties of the resin (Shen et al.). To deal with these drawbacks, coated red phosphorus was developed, which applies an organic, inorganic, or organic-inorganic coating to the red phosphorus. To create coated red phosphorus, the red phosphorus is put into an aqueous solution with another chemical. The red phosphorus and the organic or inorganic chemical react to create a precipitate on the red phosphorus, which is added to the polymer. Inorganic-organic coated red phosphorus, specifically, does not significantly impact the mechanical properties of the final material and is much more compatible with resin than organic or inorganic coated red phosphorus (Shen et al.).

Ammonium polyphosphate (APP) is also very efficient, but it absorbs a lot of moisture, turning it yellow and mechanically brittle. As Zhang et al. found, APP was an effective flame retardant for wood-fiber/PP composite materials (Zhang et al.). As the amount of APP in the compound increased, the LOI value also increased. An increase in LOI generally indicates a better flame retardant. Additionally, the study shows that when even a small amount of silica is added to the compound, it can help enhance the flame retardancy of the compound. Despite increasing the overall flame retardancy of the material, both APP and silica were shown to decrease the impact strength when added to the wood compound (Lee et al.). Zhang et al. proposed that this decrease in impact strength could be attributed to cavities that form inside the compound when processed (Zhang et al.).

Boron Flame Retardants

While boron-based flame retardants are typically not used alone, they are still an important flame-retardant category. The most commonly used boron flame retardant is zinc borate ($2ZnO \cdot 3B2O3 \cdot 3.5H2O$). This flame retardant is non-toxic and smokeless. It is commonly used with aluminum trihydroxide, mentioned in the previous section, to reduce the cost of the material. When boric acid was added to a composite, there was minimal impact on tensile strength; however, it was only tested at 1-5 wt%. The maximum elongation of a composite with boric acid was 0.02 N/mm^2 above the reference, which was seen with the 3 wt% sample (Nam et al.)

Zinc borate works by absorbing a large amount of heat and releasing water vapor to cool the polymer. According to Shen et al., for every gram of zinc borate that is decomposed, 924 J of heat is absorbed, and the water vapor produced dilutes the combustible gas in the air. To put this amount of heat in context, one gram of water takes 2259 J of heat to evaporate. Additionally, B_2O_3 is produced and coats the polymer's surface to act as an isolation film.

Organic Flame Retardants

Halogen-based additives

Halogen additives are one of the main subcategories of organic flame-retardant additives, which, in parallel to inorganics, contain carbon in their chemical backbone structure. They are also one of the most commonly used additives for almost all types of plastics (Liu et al.). Some

of the more common areas of application include furniture, mattresses, carpets, electronic devices, construction, and others (Chantler et al.). These additives most commonly contain bromine and chlorine atoms, as the bond stability of these halogens is in the optimal range for flame retardants. If the bond is too weak, like in iodine, the halogen will be released into the environment when exposed. On the other hand, if the bond is too strong, like in fluorine, the compound will fail to release the halogen radicals. However, Liu et al. mentioned that the bond strengths of the various halogen flame retardants did not quantify the strengths or bond distances of the different halogens (Liu et al.). X-ray absorption spectroscopy can be used to quantitatively measure the bond strength of the various halogen-carbon bonds of different additives (Chantler et al.). Additionally, programs that use molecular dynamics and density functional theory calculations can give accurate estimates of the bond lengths and bond energies of these carbon-halogen bonds (Mao et al; Van et al.). These quantitative measurements of the bond lengths and energies of the carbon-halogen bonds could provide insight into the optimal bond strength for these additives.

The main mechanism by which halogen additives work is by capturing free radicals released in the vapor phase during combustion. The compound releases halogen radicals when exposed to heat, which will capture the •OH and •H radicals produced in the combustion of the polymer. This yields halide gas that sits on top of the polymer and decreases the concentration of oxygen and combustible gas. Additionally, the halide gas produced is inert and denser than air, which allows it to displace oxygen away from the polymer.

The most common halogen compounds that are used are Tetrabromobisphenol A (TBBPA), 1,2‑Bis(2,4,6‑tribromophenoxy)ethane (BTBPE), and Decabromodiphenyl Ethane (DBDPE). However, DBDPE has recently been identified as a substance of very high concern by the EU (Breaking). The EU has ruled that DBDPE violates Article 57e, as it is very bioaccumulative in both the environment and in humans (Candidate). Halogen compounds are often cited as being cheap, easy to produce and process, and very effective. DBDPE has also been shown to improve the impact strength of the resulting composite (Shen et al.; Lee et al.).

Organophosphorus Flame Retardants

Another large category of organic flame retardants is organophosphorus flame retardants. Their numerous advantages have caused them to be researched extensively recently and added to most types of polymers (Liu et al.). Organophosphorus flame retardants are praised for being very efficient at flame retarding materials. Common compounds of this category include phosphines, phosphine oxides, phosphonates, and others. Specifically, 9,10-dihydro-9,10-oxa10-phosphaphenanthrene-10-oxide (DOPO) has attracted attention from researchers due to its thermal stability, oxidation resistance, and hydrolysis resistance (Liu et al.). DOPO, when added at 2 wt%, was shown to decrease tensile strength by 1.7% (Liu et al.; Sai et al.).

These flame retardants work by decomposing into PO, PO_2, and other radicals, which combine with •OH and •H radicals to stop the combustion process. Additionally, the

non-combustible gas produced reduces the concentration of oxygen, which helps inhibit combustion. During the condensed phase, the decomposition of the flame retardant yields phosphoric acid, which forms a glass layer on the polymer's surface that insulates the polymer from heat and gas. This glass layer is mechanistically similar to the previously mentioned char layer.

Inorganic-organic hybrids

Silicon

Silicon flame retardants offer many advantages, such as producing low smoke, being cheap, and being non-toxic. Additionally, the mechanical properties of polymers with silicon-based flame retardants are excellent. These mechanical benefits include high flexibility, thermal stability, being impact resistant, and hydrophobic (Shen et al.). However, Huang et al. researching P-HBPSi@GO, which is a phosphorus-containing silicon compound in a TPU matrix, showed a 21.6% reduction in tensile strength (Liu et al.; Huang et al.).

The most common silicon-based flame retardant is silicon dioxide (SiO_2). Despite having numerous mechanical advantages, silicon dioxide is inefficient and does not flame retard well. To mitigate this problem, these additives are often combined with other additives to enhance the polymer's flame retardancy. An example of a silicon-organophosphorus additive studied is 0.8 wt% SiO_2 mixed with diphenylphosphoric acid (DDP). This compound exhibited an LOI value of 29.3%, which is generally regarded as self-extinguishing.

Intumescent systems

Intumescent flame-retardant systems are somewhat unique among flame retardants because they are composed of three different components: an acid source, a carbonization agent, and a blowing agent. Generally, the ideal ratio of these three components is about 1:0.5:0.3 (Liu et al.). These compounds are often non-toxic, isolate heat and oxygen well, and suppress smoke production. Usually, three different sources are added to the polymer for each of the three components; however, this creates problems such as poor processability, poor mechanical properties, low efficiency, and nonhomogeneity. The poor mechanical properties are likely caused by the poor mixability and processability, which creates pockets of high or low concentrations of additive.

Intumescent flame retardants work by creating an intumescent carbon layer on the polymer. When the polymer combusts, the flame retardant swells up and creates the carbon layer, insulating the polymer from heat and oxygen. Figure 5 shows a diagram comparing the mechanisms of conventional and "all-in-one" flame retardants. Additionally, Figure 5 demonstrates that intumescent flame retardants create a large char layer that insulates the polymer from oxygen and heat. Unlike the char layers of other additives, intumescent flame systems generate a char layer that foams up and becomes larger.
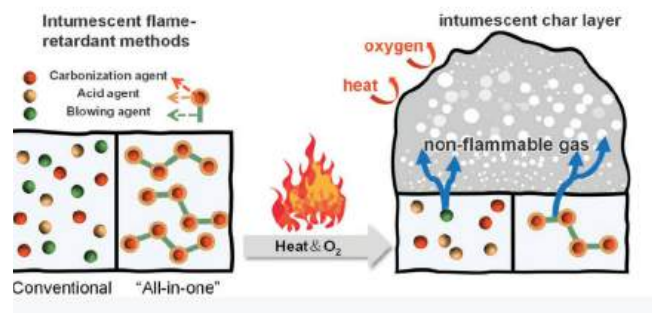
Figure 5. Diagram of the mechanism of intumescent flame-retardant additives (Liu et al.)

To help minimize the poor processability and mixability of the three sources of intumescent flame retardants, "all-in-one" intumescent flame retardants have been developed, which use the same mechanism as intumescent flame-retardant systems in one molecule. As shown in Figure 5 above, 'all-in-one' intumescent additives can disperse more evenly within the polymer. They are an improvement over conventional intumescent additives, which may have pockets of varying concentrations of the components, which would negatively impact the efficiency of the additive. One of the higher achieving all-in-one systems researched recently was a piperazine-modified ammonium polyphosphate molecule (PAz-APP). This PAz-APP system achieved an LOI value of 31.2% and a UL-94 rating of V-0 (Liu et al.). An LOI value of 31.2% is above the 28% accepted threshold for self-extinguishing materials.

Nano-Fillers

Nano-filler flame retardants are a relatively new flame-retardant group. Previously, micro materials were used, but they were too inefficient because of a low surface area-to-weight ratio. With new nano-materials that were developed, the surface area to weight ratio has greatly increased, which now allows nano-fillers to be a viable flame retardant.

Nano-filler flame retardants help enhance the formation of carbon, which reduces the heat release rate of the polymer. This reduces the ignition time of the polymer and the amount of combustible gas produced.

Nano filler flame retardants offer better heat stability, low water absorption, and superior mechanical properties to other conventional flame retardants. However, this class of flame retardants is still being researched, and widespread adoption is yet to be achieved. As this technology is very new, nano-fillers are currently much more expensive than other conventional flame retardants. Nano-fillers can be used in synergy with other flame retardants to enhance the mechanical properties of the polymer without decreasing flame retardancy (Liu et al.).

Layered Silicate Nanocomposites

Polymer/layered silicate nanocomposites (PLSN) are a relatively new flame retardant that have shown promising results. PLSNs offer good thermal stability, mechanical properties, and flame retardancy (Liu et al.). There are various types of layered silicates that are used as flame

retardants, which include montmorillonite (MMT), mica, fluorohectorite, bentonite, saponite, and hectorite, etc (Liu et al.).

PLSNs work by creating a char layer on the surface of the polymer and by increasing the viscosity of the melted polymer. The increased viscosity of the melted polymer reduces melt dripping, which prevents the spread of the flame to the environment.

Graphene-based

Like PLSNs, graphene-based nanocomposites are relatively new flame retardants and have gained significant attention because of their promising advantages. These advantages include high specific surface area, high electrical conductivity, good mechanical properties, and good thermal conductivity (Liu et al.).

Graphene-based flame retardants work based on two mechanisms: the creation of a tortuous path and the creation of a char layer. The creation of a tortuous path is unique to graphene-based nanocomposites and works by changing the diffusion of pyrolysis products(Liu et al.). Figure 6 shows the effect of the tortuous path on the combustion of the polymer. Instead of traveling directly to the surface of the polymer to be combusted, the combustible volatiles have to take indirect paths to the surface of the polymer. Without the blockades from the graphene flame retardants, the combustible volatiles would take a more direct path out of the polymer and react faster with the outside heat and oxygen. Additionally, the heat and oxygen from the environment cannot take a direct path into the polymer because of the graphene additive in the way. This lengthened path helps slow the combustion of the polymer. As discussed previously, the creation of a char layer helps insulate the polymer from heat and gas.
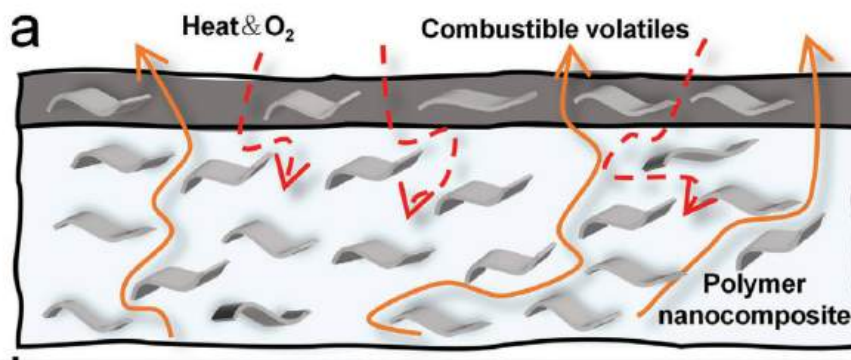
Figure 6. Diagram of how the tortuous path mechanism of graphene-based nano-fillers works (Liu et al.)

Conclusion

Modern polymers have gained substantial popularity over the last century and are now used in almost every aspect of life. It is hard to understate their utility and popularity in society today. However, the chemical structure of polymers, being mainly carbon and hydrogen, causes them to be very flammable and has contributed to fire disasters around the world. While researchers have developed commercially available and inherently flame-retardant polymers, these polymers are often much more expensive than current consumer-level polymers. To

address this issue of combustion, flame-retardant additives have been developed and researched with varying results. One common way to quantify the effectiveness of flame-retardant additives is to use tests such as the LOI and the UL-94 tests.

There are many different types of flame-retardant additives that are used today, ranging from inorganic metal hydroxides to cutting-edge nano-fillers. Each additive has its advantages and disadvantages. Notable additives include halogen, metal hydroxides, intumescent, and nano-filler additives. Halogen additives are currently the most used, but are toxic to both humans and the environment. Regulatory bodies have placed strict regulations on many halogen-based additives, so other additives are gaining popularity. Metal hydroxides are non-toxic; however, they significantly negatively impact the mechanical properties of the polymer composite. Intumescent and nano-filler additives are newer flame-retardant additives that yield promising results. Intumescent additives use a char-forming mechanism, while nano-fillers create torsion paths inside the polymer. Granted, intumescent systems have poor mixability, and nano-fillers fail to delay ignition time.

Overall, "all-in-one" intumescent flame retardants are promising additives that are effective at flame retarding materials. Unlike traditional intumescent flame retardants, "all-in-one" systems avoid the poor mixability of traditional intumescent systems. Additionally, with the unique foaming char mechanism, it can surpass other char-generating flame retardants. Another outstanding group of flame retardants is the metal-hydroxide group because of its low cost and non-toxicity. However, metal-hydroxide additives should only be used when the mechanical properties of the final composite are insignificant.

Future research into flame-retardant additives could explore synergies of different combinations of additives in polymers. For example, a study analyzing the properties of carbon-nano fillers with magnesium hydroxide would be interesting. Current review papers on flame-retardant additives usually describe the effects of additives on polymers using qualitative analysis rather than quantitative analysis (Shen et al.). The papers might claim that a certain additive has high flame-retardant efficiency but will not mention any quantitative measurement to reinforce the claim. Nano-filler flame retardants show potential for widespread use, but their current high cost will likely keep them from being incorporated into current polymers. Moreover, flame-retardant additives for polymers have the potential to be applied to other composites; however, the compatibility with other materials needs to be researched before widespread implementation of polymer additives into other materials is adopted.

Unequal Partnership or Mutual Gain? The Hungary–China Economic Equation
Who Gains More? Evaluating the Hungary–China Economic Partnership
By Matte Andrew Keller

Abstract

As global power dynamics realign, small and mid-sized countries like Hungary find themselves in increasingly influential roles. Once an insignificant player in global politics, it now serves as a pivotal chess piece in the growing competition between the United States and China. Hungary's journey, from the Central Powers in World War I, to Axis alignment in World War II, to decades behind the Iron Curtain, has shaped a country's awareness of the cost of dependency and the allure of sovereignty. Since the fall of Communism in 1989, Hungary has positioned itself as a bridge between East and West, joining NATO in 1999 and the European Union in 2004. But recent tensions with Brussels and Washington have led Budapest to look further east once again, this time, toward China.

This pivot is strategic. From under €100 million a year for most of the 2010s, Chinese FDI in Hungary surged after 2021, reaching nearly €3 billion in 2023 and continuing to climb as Hungary became China's top investment destination in Europe (MERICS). In fact, according to some reports last year Hungary attracted 44% of all Chinese investment in Europe, more than any other EU country (MERICS). Behind those numbers are major projects from Chinese giants like CATL, which is constructing a €7.3 billion battery plant in Debrecen (CEIAS*)*, and BYD, which plans to open a €5 billion electric vehicle factory in Szeged (Central European Times).

These projects offer Hungary jobs, capital, and global visibility, but also raise fundamental questions. What does Hungary gain in the long run? And is the deal as balanced as it seems?

China's European Strategy and Hungary's Role in It

For China, investing in Hungary is a smart, strategic play. Hungary, as an EU member, offers China an internal production base that can bypass tariffs and regulatory hurdles facing goods made in mainland China (OSW). Products manufactured in Hungary enjoy smoother entry into the 27 member European single market, giving Chinese companies like CATL and BYD a powerful competitive edge (MERICS). Cheaper logistics, fewer customs headaches, and proximity to European consumers make Hungary a dream location for export-oriented growth (MERICS).

But this goes beyond economics. Hungary also is supporting China's political interests in a way that is unique in the EU. For example, Hungary has blocked EU resolutions criticising Beijing over Hong Kong and Xinjiang (Reuters). Hungary also has consistently sought bilateral agreements with China that limit Brussels oversight (Reuters). For China, Hungary is not only a place to do business, it is a way into European decision making.

This model is part of China's larger strategy: build leverage through infrastructure, diplomacy, and capital. Economic engagement opens doors to political influence. Hungary with its increasing skepticism of Western institutions, has proved a willing partner.

In short, China is playing a long term game. It is using Hungary not just to enjoy a short term economic gain in the EU market, but to test the political benefits it can receive through this relationship for, perhaps, decades to come.

Hungary's Gamble on Growth

Hungary's embrace of the Chinese capital is not without rationale. The country has long struggled with regional inequality, and areas like eastern Hungary have lagged in industrial development (OECD). Projects like CATL and BYD promise to revitalize cities like Debrecen and Szeged, bringing thousands of new jobs and modern infrastructure. Hungarian officials have highlighted waves of Chinese projects as drivers of economic growth, emphasizing these projects enhance regional development and signal Hungary's openness to global partnerships (OSW).

However, while the job creation numbers are real, their quality and sustainability are less convincing. Many of these jobs are in construction or basic logistics, roles that are temporary, low skill, and often filled by subcontractors or foreign workers (Central European Times). The promise of long term employment, upskilling, and leadership opportunities for Hungarian citizens remains unproven.

Moreover, Hungarian firms have largely been excluded from high-value segments of the supply chain (MERICS). Core components, overall company structure, and decision making remain in the hands of Chinese parent companies (Friedrich-Ebert-Stiftung). This limits the spillover effects for the local economy. Hungary is hosting the factory, but not necessarily participating in the innovation.

Without deliberate policy interventions, like technology sharing agreements or local supplier mandates, Hungary risks becoming a service platform, not a strategic partner.

The Case for a Smarter, Stronger Bargain

Hungary's openness to investment is not the issue. The challenge lies in the terms of engagement. In seeking capital, Budapest should not lose sight of what the capital is meant to achieve. True economic development means more than GDP growth; it means building domestic capacity, skills, and competitiveness.

There are clear achievable steps Hungary can take to improve the value of these partnerships. First enforce stronger local content requirements. Guaranteeing that 60–70% of jobs go to Hungarian workers would help ensure that employment opportunities strengthen the local workforce and build long-term economic stability. Requiring that 30% of contracts go to Hungarian-owned companies would further embed foreign investment into the domestic economy, supporting local businesses and reducing dependency on external supply chains.

Second, Hungary should negotiate for deeper integration of local universities and technical schools in these projects. Funding research partnerships, innovation hubs, and vocational training programs would ensure investments do not just employ people, they empower them.

Third, Hungarian authorities should develop a long term FDI strategy that aligns with

national priorities in energy, transportation, and digital infrastructure. Investment should not be reactive, but strategic, shaped by national goals rather than external convenience. For example, Hungary should consider requiring any investment proposal to demonstrate how the energy usage for the site will either be net-zero, or even net-positive, to Hungary's national energy system.

None of this requires rejecting Chinese investment. It simply requires treating Hungary not as a venue for foreign ambitions, but as an equal stakeholder in its own economic future.

What China Stands to Gain from Reciprocity

China, too, has a stake in a more balanced relationship. As it expands across Europe, Beijing faces increasing scrutiny. The EU has opened an anti-subsidy investigation into Chinese EV imports, including BYD, reflecting rising concerns over whether Chinese firms are playing by Europe's competition and subsidy rules. (Università Bocconi*)*

To strengthen its standing, China should prioritize transparency, compliance, and cooperation with local actors. Supporting Hungarian small and medium sized enterprises, investing in joint research and development, and committing to local workforce development would not only reduce political friction, but build trust. The more China is seen as a fair partner, the more sustainable its footprint becomes.

Hungary is not the only European country interested in Chinese capital, but it is the one that has given China the most favorable terms. If China hopes to scale this model across the EU, making Hungary a showcase of mutual benefit, not exploitation, will be key.

Conclusion: From Deal to Dialogue

Hungary and China are redefining what a 21st century economic partnership looks like. At its best, this relationship shows how nations of vastly different sizes can collaborate for mutual gain. But without recalibration, the risk is clear: Hungary may gain in the short term, while China secures the long term advantage.

A better deal is possible. It does not require confrontation, it requires dialogue. Hungary must negotiate not just for jobs and factories, but for participation, innovation, and sovereignty. And China must recognize that its power lies not in dominance, but in the trust it builds as a reliable partner.

In an uncertain global order, small countries like Hungary matter more than ever. But that influence is only meaningful if paired with a clear-eyed sense of national interest. With the right terms, Hungary can remain open to global investment while shaping its own future, not merely hosting someone else's.

This is not just smart politics. It's sustainable economics, for both nations.

Works Cited

"China's Controversial €7.3 Billion EV Battery Investment: A Game Changer in China-Hungary Relations." CEIAS, Nov. 2023, www.ceias.eu/chinas-controversial-e7-3-billion-ev-battery-investment-a-game-changer-in-china-hungary-relations/.

"Hungarians Increasingly Displaced by Guest Workers - Report." Central European Times, Jan 2024, www.centraleuropeantimes.com/hungarians-increasingly-displaced-by-guest-workers-report/.

Chalmers, John. "Hungary Blocks EU Statement Criticising China over Hong Kong, Diplomats Say." Reuters, Apr. 2021, www.reuters.com/world/asia-pacific/hungary-blocks-eu-statement-criticising-china-over-hong-kong-diplomats-say-2021-04-16/.

Kratz, Agatha. "Dwindling Investments Become More Concentrated - Chinese FDI in Europe: 2023 Update." MERICS, June 2024, merics.org/en/report/dwindling-investments-become-more-concentrated-chinese-fdi-europe-2023-update.

"Chinese vehicle maker BYD plans plant in Szeged" Central European Times." Dec. 2023, centraleuropeantimes.com/chinese-vehicle-maker-byd-plans-plant-in-szeged/.

Gizińska, Ilona, and Paulina Uznańska. "China's European Bridgehead. Hungary's Dangerous Relationship with Beijing." OSW, Apr. 2024, www.osw.waw.pl/en/publikacje/osw-commentary/2024-04-12/chinas-european-bridgehead-hungarys-dangerous-relationship/.

"OECD Regional Outlook 2023: Hungary." OECD, Aug. 2023, www.oecd.org/en/publications/oecd-regional-outlook-2023-country-notes_be47635c-en/hungary_88363f9e-en.html.

Czirfusz, Márton. "The Battery Boom In Hungary: Companies of The Value Chain, Outlook for Workers and Trade Unions." Friedrich-Ebert-Stiftung, 2023, library.fes.de/pdf-files/bueros/budapest/20101.pdf.

Hu, Weinian, and Daniel Gros. "THE BEV ANTI-SUBSIDY INVESTIGATION OF THE COMMISSION INDUSTRIAL POLICY AND GEOPOLITICS IN DISGUISE?" Universita Bocconi, Apr. 2024, iep.unibocconi.eu/sites/default/files/media/attach/Hu_Gros_BEVanti-subsidy.pdf.

The Strategies Used in Marketing that Currently Impact Adolescent Oral Hygiene Habits: A Literature Review By Anya Mantena

ABSTRACT

Adolescent oral health remains a major public health concern in the United States. More than half of adolescents have experienced dental caries, and a notable share have untreated disease. This literature review examines how advertisements, influencer posts, and sponsorships on social media influence oral hygiene among adolescents aged 11–18. Using structured searches across PubMed, Google Scholar, and related databases, fourteen key peer-reviewed studies were selected from an initial pool of over 500 results based on relevance, quality, and focus on adolescent oral health. Findings indicate that TikTok, Instagram, and YouTube are major channels for both oral health education and commercial promotion. Visual and interactive formats increase engagement but can also spread misinformation. Influencers wield significant persuasive power and can blur the boundary between education and promotion. Sponsorships and institutional programs offer structured, evidence-based content but remain under-evaluated for their specific digital impacts. Strengthening media literacy, ethical marketing standards, and research rigor is essential to ensure adolescents receive accurate and trustworthy oral health information.

INTRODUCTION

Maintaining oral health is a critical public health priority for adolescents. In nationally representative data, approximately half of U.S. adolescents aged 12–19 have had dental caries, and more than one in ten have untreated disease (Figure 1).
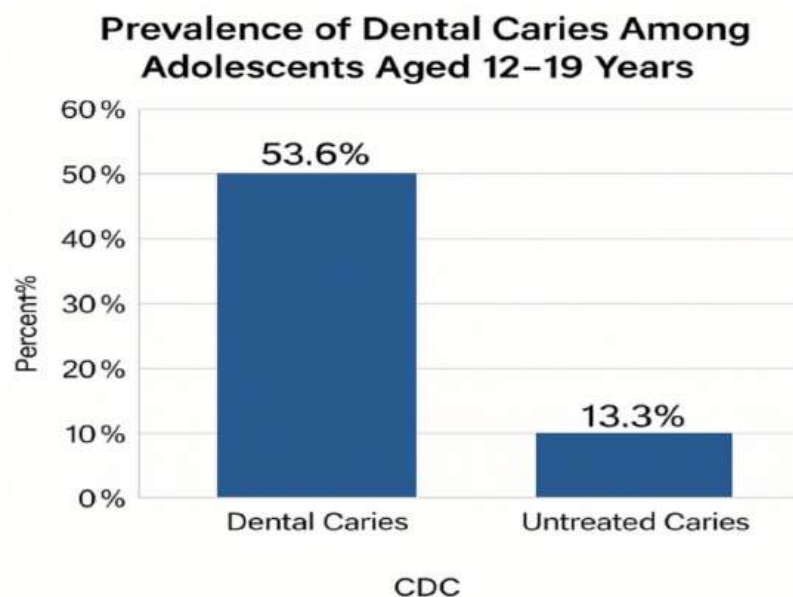

Figure 1. Prevalence of dental caries among U.S. adolescents aged 12–19 years, based on NHANES 2015–2016 data.[1] Higher rates of untreated caries highlight persistent oral-health disparities among U.S. adolescents.

Untreated oral conditions are linked with pain, infection, difficulty eating and speaking, reduced quality of life, and tooth loss; tooth loss itself is associated with functional, psychosocial, and nutritional consequences.[2,3] Excess free-sugar intake, especially from sugar-sweetened beverages, is a well-established risk factor for dental caries in children and adolescents.[4]

Concurrently, social media use is ubiquitous in this age group: Most U.S. adolescents use social media, and nearly half report being online "almost constantly".[5] Platforms such as TikTok, Instagram, and YouTube shape perceptions and health-related decisions. This digital ecosystem presents opportunities for timely, interactive oral-health education and risks of commercial bias, aesthetic-driven messaging, and misinformation.

This review synthesizes evidence on three dominant marketing strategies used on social media that may influence adolescent oral hygiene behaviors such as commercial advertising, influencer promotions, and organizational sponsorships; highlighting both educational potential and risks while outlining gaps for future research.

METHODS AND MATERIALS

Search Strategy

A structured literature search was conducted between February and June 2025 in PubMed and Google Scholar, complemented by targeted searches of dental/public-health journals and organizational reports. Search terms included: *social media marketing oral hygiene adolescents*, *influencer marketing oral health teens*, *social media advertising dental care teenagers*, *sponsorships oral health campaigns youth*, and *social media adolescent oral health United States*.

Screening and Eligibility

Approximately 580 records were identified. After removing duplicates and screening titles/abstracts for relevance to social-media marketing and adolescent oral health in the U.S., non-English publications and items prior to 2020 were excluded to ensure recency.

Inclusion criteria: (i) peer-reviewed empirical studies or systematic reviews (2020–2025); (ii) populations including adolescents aged 11–18 years; (iii) analysis of social-media advertising, influencer activity, or sponsorship campaigns related to oral health.

Exclusion criteria: studies not focused on oral health behavior; editorials/commentaries without data; studies without U.S. relevance.

Selection and Data Extraction

Forty-two full texts were assessed for eligibility. Fourteen studies were selected based on methodological quality, diversity of platforms, and contribution to understanding behavioral

outcomes. Screening and data extraction were performed by a single reviewer (A.M.); final inclusion decisions and extracted data tables were independently verified by a supervising mentor for accuracy (10% random subsample), with discrepancies resolved by discussion.

RESULTS

Commercial Advertising on Social Media

Commercial dental advertising targeting adolescents commonly emphasizes rapid cosmetic outcomes (e.g., whitening) and aesthetics (Figure 2).
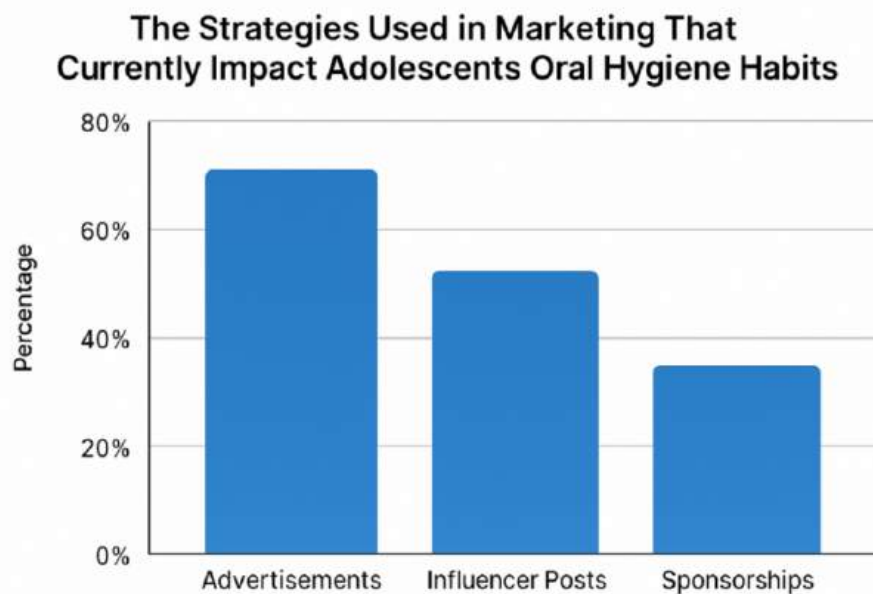
Figure 2. Examples of social-media dental marketing content emphasizing whitening-focused and cosmetic messaging targeted toward adolescents.[6] Digital marketing can influence adolescent oral-hygiene behaviors positively or negatively, depending on content credibility and message framing.

Systematic evidence indicates that social media can influence oral-health knowledge and behaviors, yet promotional content often outpaces evidence-based education.[6] Platform-native formats (short video, motion graphics, filters) are designed to maximize engagement, which can enhance recall but may also promote product-centric claims without adequate substantiation.[6] In contrast, structured educational campaigns that blend accurate messaging with engaging formats have shown improvements in short-term self-reported behaviors such as brushing frequency or plaque indices in youth and orthodontic patients.[7,8]

Influencer Marketing and Adolescent Oral Hygiene

Influencer endorsements, both by lifestyle creators and dental professionals, are highly salient to adolescents' purchasing decisions and perceived norms.[5] Analysis of Instagram content

shows that influencer cues (credibility, aesthetics, and social proof) significantly shape toothpaste selection and related attitudes.[7] Randomized evidence in young orthodontic patients found that integrating Instagram with chairside instructions improved oral-hygiene outcomes compared with instructions alone.[8] However, content quality varies; unverified trends (e.g., charcoal powders or DIY whitening) risk enamel abrasion and sensitivity and may lack fluoride benefits essential for caries prevention.[9-11]

Organizational Sponsorships and Public-Health Campaigns

Sponsorships by dental organizations, nonprofits, and corporate social-responsibility partners can deliver standardized curricula and credible messaging at scale. For example, CATCH My Breath (an evidence-based, peer-led program for grades 5–12) has documented prevention outcomes and includes materials addressing oral-health risks related to vaping.[12,13] Partnerships by payers and nonprofits (e.g., DentaQuest with America's ToothFairy and CATCH) illustrate how sponsorships can amplify reach across school and community settings.[13,14] Despite these advantages, many evaluations aggregate overall health outcomes and rarely isolate the incremental effects of social-media engagement, signaling a need for more granular digital-evaluation designs.

DISCUSSION

Across modalities, social media exerts a measurable influence on adolescents' oral-health knowledge, attitudes, and behaviors.[6–8] Advertising efficiently captures attention but often prioritizes aesthetics and rapid results over prevention messaging. Influencer marketing is a paradox: professional or evidence-guided creators can improve adherence (e.g., plaque control), whereas non-expert trends have propagated abrasive or fluoride-deficient practices associated with enamel wear and sensitivity.[9–11] Sponsorships and institutional programs bring curricular rigor, health-behavior theory, and implementation supports; early evidence of effectiveness in adjacent outcomes (e.g., vaping prevention) suggests potential spillovers to oral-health literacy, but digital-specific impacts remain under-reported.[12–14] Two cross-cutting needs emerge. First, clearer platform-level and professional guidelines are warranted to ensure that dental promotions disclose commercial interests, cite evidence, and avoid unsafe claims. Second, adolescent media-literacy training that emphasizes how to evaluate digital health claims could buffer against misinformation and aesthetic-only appeals.

CONCLUSION

Social media is deeply embedded in U.S. adolescents' daily lives, with many reporting near-constant online activity.[5] This creates both an unprecedented channel for evidence-based oral-health education and a pathway for misinformation and product-driven messaging. Current evidence shows that interactive, platform-native strategies and professionalized influencer content can improve short-term oral-hygiene behaviors[7,8], whereas unverified trends (e.g., charcoal abrasives, DIY whitening) may pose risks to enamel and caries prevention.[9–11]

Sponsorship-backed programs demonstrate scalable, credible approaches, yet their *digital* contribution to outcomes requires more rigorous evaluation.[12–14] Future research should prioritize longitudinal designs linking social-media exposures to behavioral adherence and caries outcomes, and policymakers and dental associations should institute transparent, enforceable standards for online dental advertising and influencer disclosures.

ACKNOWLEDGEMENTS

CONFLICT OF INTERESTS

The author declares no conflicts of interest.

Works Cited

1. Fleming E, Afful J. Prevalence of Total and Untreated Dental Caries in Youth: United States, 2015–2016. NCHS Data Brief No. 307. Hyattsville, MD: National Center for Health Statistics; 2018. Available from: https://www.cdc.gov/nchs/products/databriefs/db307.htm (Accessed 2025-06-25).

2. Centers for Disease Control and Prevention. Oral Health: Facts. 2024. Available from: https://www.cdc.gov/oral-health/data-research/facts-stats/index.html (Accessed 2025-06-25).

3. Centers for Disease Control and Prevention. About Tooth Loss. 2024 May 15. Available from: https://www.cdc.gov/oral-health/about/about-tooth-loss.html (Accessed 2025-06-25).

4. Chi DL. Added Sugar and Dental Caries in Children. Pediatr Clin North Am. 2018;65(5):991-1001. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6242348/ (Accessed 2025-06-25).

5. Pew Research Center. Teens, Social Media and Technology 2023. 2023 Dec 11. Available from: https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/ (Accessed 2025-06-25).

6. Alharthi SS, Alghamdi FA, Alfaqeeh SA, Alsubait SS, Alolayan AA, Alqahtani AM, et al. The impact of social media on oral health: A systematic review. PLoS One. 2024;19(4):e0296102. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0296102 (Accessed 2025-06-25).

7. Abdullah NO, Abu-Hammad OA. Decoding the influence of Instagram content on consumer behavior in toothpaste selection: Implications for oral health and marketing strategies. Front Oral Health. 2024;5:1420500. Available from: https://www.frontiersin.org/articles/10.3389/froh.2024.1420500/full (Accessed 2025-06-25).

8. Scribante A, Gallo S, Bertino K, Meles S, Gandini P, Sfondrini MF. The effect of chairside verbal instructions matched with Instagram social media on oral hygiene of young orthodontic patients: A randomized clinical trial. Appl Sci. 2021;11(2):706. Available from: https://www.mdpi.com/2076-3417/11/2/706 (Accessed 2025-06-25).

9. Zoller MJ, Schonhals M, Schmidt S, Naumova EA, Arnold WH. Relative dentin and enamel abrasivity of charcoal-containing toothpastes. Sci Rep. 2022;12:6713. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10099862/ (Accessed 2025-06-25).

10. Yaghini J, Gholami M, Aslani H, Rakhshan V. Effect of two types of charcoal toothpaste on enamel surface roughness: An in vitro study. Dent Res J (Isfahan). 2023;20:78. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10630542/ (Accessed 2025-06-25).

11. Lemke DDS. Charcoal toothpastes pose risks with few benefits. J Mich Dent Assoc. 2022;104(5):38-41. Available from:

https://commons.ada.org/journalmichigandentalassociation/vol104/iss5/7/ (Accessed 2025-06-25).

12. Burchfield KM, Collins C, Erwin PC, Greene SB, James HJ. The CATCH My Breath vaping prevention curriculum: Implementation and evaluation in central Appalachia. Prev Med Rep. 2024. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11403683/ (Accessed 2025-06-25).

13. DentaQuest. Community giving and partnerships (CATCH My Breath collaboration). 2025. Available from: https://www.dentaquest.com/en/about-us/community-giving (Accessed 2025-06-25).

14. America's ToothFairy. America's ToothFairy and DentaQuest team up to fight teen vaping. News release; 2022 Feb 1. Available from: https://www.americastoothfairy.org/press-releases/americas-toothfairy-and-dentaquest-team-up-to-fight-teen-vaping-during-national-childrens-dental-health-month (Accessed 2025-06-25).

American vs European Growth and the Shortcomings of GDP By Ethan Cao

The United States' unprecedented economic growth during the Gilded Age and onwards has completely revolutionized the world, drastically improving the quality of life to what we know today. The primary statistic used to determine the immense success of the United States is Gross Domestic Product (GDP), but many of the factors of the standard of living are ignored by GDP. This research paper aims to delve into other metrics that support GDP in painting a clearer picture of the United States' true growth compared to Europe.

The Gilded Age was a time period in the United States of immense industrialization, modernization, and urban expansion after the Civil War and Reconstruction (Digital history, 2025). The period marked the United States' emergence as a global superpower driven by railroads, steel, oil, and innovative finance. Additionally, an influx of millions of immigrants from Europe and Asia in search of opportunity contributed to the growth of cities, but also to overcrowding, rising health and sanitary issues, and poverty. The divisions between social classes deepened, leading to calls for social reform and the adoption of many laws and regulations. The innovations of the Gilded Age, along with the gradual decrease in their costs, greatly improved life expectancy, productivity, and the overall quality of leisure for society. These innovations came through many different aspects of life.

Although in general, GDP is the main metric used when discussing a country's economic status and growth, it does not fully encapsulate all of the measures that are related to a country's success. For example, the quality of things like housing, food, and healthcare all improved over time, but due to the relatively stagnant monetary cost of all of these commodities, GDP in the US did not increase as much as it should have to fully reflect the actual impact that the improvement in these commodities had (Gordon, 2017). Additionally, life expectancy and quality of leisure time both greatly improved in the late 19th century and early 20th century, allowing people much more freedom and greatly impacting the quality of life overall (Gordon, 2017; Dattani et al., 2023). However, due to the non-transactional nature and also not being a tradable good, these factors are virtually completely ignored by GDP statistics, completely understating the vast amount of improvement actually achieved during the period around the Gilded Age.

One of the largest changes was in food. Refrigeration and other new preservation methods, such as the Mason jar, allowed people to store food safely for longer periods and also enjoy a greater variety in their diets, moving beyond the "hogs and hominy" diet, which was considered traditional and common at the time (Gordon, 2017). Additionally, the influx of immigration from other countries also helped in adding variety to American diets (Gordon, 2017). Some examples include the introduction of pasta by Italians and new meat preparation techniques by Germans. Access to food expanded at the same time with the growth of supermarkets and mail-order catalogs from Montgomery Ward and Sears Roebuck, but this new transition to more industrialized food preparation also caused many problems, including the risk of food safety (Gordon, 2017). Many of these issues were highlighted in Upton Sinclair's The Jungle (Sinclair, 2021), a book that muckraked the American food industry. While the intrinsic

value of the food vastly improved, the material cost did not change much. The nutritional content and variety drastically increased, but its involvement in transactions and the economy remained relatively stagnant.

Another economic sector that rapidly grew was housing. Whereas previously, many people lived in rural homes that lacked many of the modern utilities we consider vital today, the invention of electricity, plumbing, and central heating vastly improved the quality of housing (Gordon, 2017). All of these new inventions completely revamped housing as well as the value of the new houses that included all of these utilities. Another improvement came in the form of more space per person as families increasingly lived with fewer people per room, and many of the new appliances invented, like the dishwasher and electric iron, reduced household labor (Gordon, 2017). In many small towns, people of different socioeconomic statuses lived close together, with the possibility of having neighbors far richer or poorer than oneself, creating a sense of social equality. These improvements in living conditions definitely increased the cost of housing, but not by the three times value measured of the upgrade in housing (Gordon, 2017). Furthermore, the quality of staying at home rocketed upwards. Before, the hours after dark could not be spent meaningfully due to the lack of light, but the new electricity allowed much more to be done (Gordon, 2017). This revolutionary innovation completely revamped how people spent time, without being captured within the metrics of GDP. Additionally, access to clean water and plumbing drastically improved overall life expectancy, preventing disease and other factors that led to early death.

Another factor that allowed people to make the change from rural farms to urban centers were the innovations in transportation. These innovations changed where and also how people lived. Railroads initially connected cities with fast, reliable travel, while streetcars shaped commute within the cities (Gordon, 2017). These trains also allowed the transportation of food across large distances with the help of refrigeration to help with the variety in diets that came over time (Gordon, 2017). Later on, the invention of the automobile extended mobility to even rural Americans, providing access to stores, jobs, and services that were almost unknown before then (Gordon, 2017). The most well-known form of this was Henry Ford's Model T, which made cars affordable even for ordinary families, while also reducing the dependence on horses, which caused sanitation and disease problems in urban centers. This lessened dependence on horses also saved valuable resources through space, food, and jobs that were previously required to maintain the horses. Quickened travel meant that productivity increased, but as demonstrated previously, productivity itself is something difficult to measure through GDP. Time that was before spent on horses and wagons could now be spent working or tending to family.

Separate from housing and transportation were the inventions in communication and entertainment. These new technologies added a new dimension to daily life. The telegraph allowed long-distance communication and also supported the railroad system, while the telephone spread rapidly, becoming a uniquely American innovation that connected households and communities (Gordon, 2017). The options for entertainment also multiplied, with phonographs, radios, and motion pictures providing new forms of leisure. Newspapers and

magazines also created new ways for information and culture to be brought into homes. These forms of leisure allowed people to spend their leisure time in ways that were unheard of before, with the ability to stay interconnected with one another across communities. The newfound speed at which information could be transmitted greatly allowed the improvement in global interconnectivity (Gordon, 2017). GDP and economic success could not capture the magnitude of this new way of communication.

As people urbanized and moved from rural homes to the city, disease started to become a concern because of the newfound close proximity of people and their living quarters. The rapid urbanization lowered life expectancy since people were exposed to overcrowded housing, pollution, and limited access to fresh food and air (Gordon, 2017). Industrialization also introduced new workplace hazards and public health concerns. Many changes were needed to help reverse the effects of the rapid urbanization, and these came in the form of revamped hospitals that were able to provide more for citizens without containing the previous perils like unhygienic tools and practices (Gordon, 2017). New inventions like antiseptics and penicillin also created methods to prevent diseases that were rampant before (Gordon, 2017). The breakthroughs in germ theory that replaced homeopathy led to new realizations in the transmission of diseases, as well as how to prevent them (Gordon, 2017). The creation of the Food and Drug Administration helped prevent the spread of adulterated foods that were tampered with as well as creating the requirement for doctors to prescribe drugs created more stability and trust in medicines and foods that were bought (Gordon, 2017). All of these new methods to protect people from disease increased life expectancy, and massively decreased the rates of infant mortality, one of the major causes of death (Gordon, 2017). While these new drugs and breakthroughs in medicine created a new source of spending in healthcare, its other aspect in improving overall health is entirely missed by GDP.

During this time period, the overall economy changed greatly through the newfound involvement of the general population. Whereas previously, transactions did not occur very frequently, the new changes in creating the mail-order catalog, as well as the new inventions that seemed almost like a necessity to have, combined with the easy access to items that previously seemed almost unobtainable due to the invention of large-scale transportation like railroads allowed a lot more economic activity in the United States, while also increasing the overall wealth of the population (Gordon, 2017). This interconnectedness fostered shared growth that is missing from GDP metrics, because of the intrinsic value in communication and connections that is separate from monetary transactions.

All of these innovations greatly improved the productivity of people's time, as many more things could be done when compared to before. For example, larger distances could be covered, entertainment quality had improved, and nutrition and quality of food also increased. None of these are measured in GDP, as they are part of the inherent value of the item itself, rather than the marketed value. Additionally, quality improves over time as a result of new novelties, but due to the widespread acceptance of the new inventions, the actual price does not truly increase (Gordon, 2017). In reality, the value associated with the same price drastically

increases, but none of the GDP metrics clearly show this growth in value. Another example is general health. With the new breakthroughs in medical technology, life expectancy as well as infant mortality were both greatly improved (Gordon, 2017; Dattani, 2023). However, while the increased purchasing of medicines and drugs was included in GDP, the very value of being able to improve the overall health of the population is not demonstrated in GDP. The lack of information that GDP provides concerning the success of the United States means that all of the changes made were not captured in the true growth during the time, but rather projected as solely monetary growth. The shortcomings of GDP demonstrate that the reality of the US's growth was nearly threefold what most GDP statistics show, leaving a lot of the revolutionary improvements completely understated (Gordon, 2017). Additionally, the impact of the growth on people as they transitioned into urban cities even greater magnifies the supposed growth shown in GDP, as this newfound accessibility to what was at the time cutting-edge technology, as well as interconnectivity between the population, allowed much more improvement in the quality of society and the lives of people.

To put into perspective the information missing from GDP metrics, some graphs have been shown below:
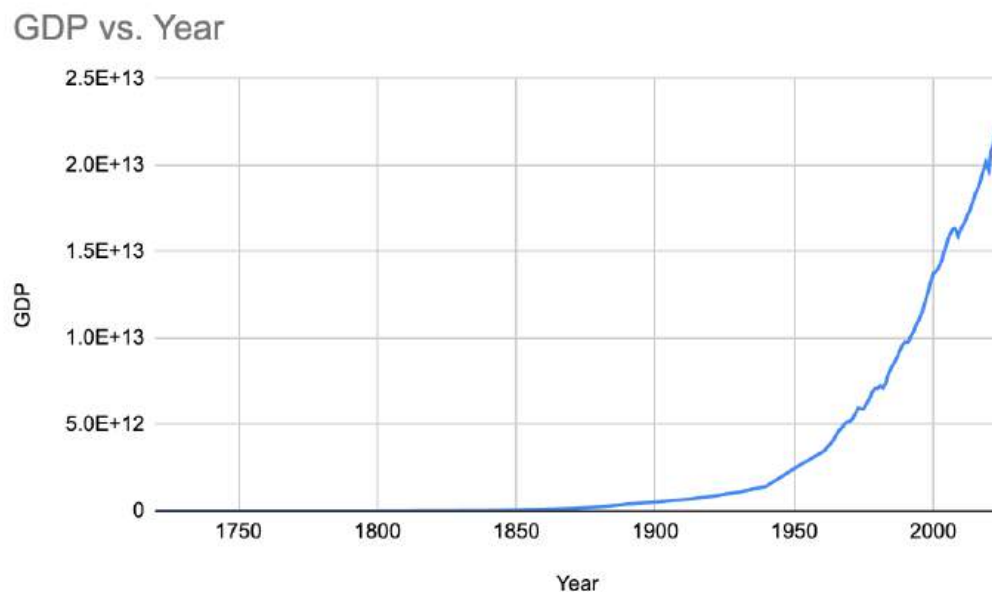
Graph of the GDP of the United States since around its inception

As shown, the Gilded Age was the beginning of the massive economic growth of the United States (Colonial and pre-federal statistics, 2025; World bank open data, 2025; Measuring worth - graphs, 2025; Rosenbloom, 2018; Mancall et al., 2025). However, when looking at total

Period	Total Factor Productivity (Average Annual Growth Rate)	Main Sources of Growth	Change in Life Expectancy at Birth (Years per Decade)
1870 to 1900	~ 1.5% to 2%	Transportation, communications, trade, business organization	1.3
1900 to 1920	~ 1%		3.2
1920s	~ 2%	Electricity, internal combustion engines, chemicals, telecommunications	5.6
1930s	~ 3%		3.2
1940s	~ 2.5%		5.3
1950 to 1973	~ 2%	Widespread	1.4
1973 to 1990	< 1%		2.4
1990s	> 1%	Information technology	1.7
2000s	~ 1.5%		1.4
1870 to 2010	~ 1.6% to 1.8%		2.3
1950 to 2010	~ 1.2% to 1.5%		1.8

Sources: Field (2012), Gordon (2010), Carter et al. (2006), Center for Disease Control and Prevention (http://www.cdc.gov/nchs/data/dvs/deaths_2010_release.pdf).

productivity and life expectancy, it becomes clear just how much, in reality, is left out by GDP. Here, it is clear that during the years during and following the Gilded Age, total factor productivity rose massively, following the many innovations of the time (Shackleton, 2013). However, the slow rise in the GDP graph barely seems to take off at this time.

Life expectancy

Period life expectancy[1] is the number of years the average person born in a certain year would live if they experienced the same chances of dying at each age as people did that year.

Data source: Riley (2005); Zijdeman et al. (2015); HMD (2024); UN WPP (2024)   OurWorldinData.org/life-expectancy | CC BY

1. Period life expectancy Period life expectancy is a metric that summarizes death rates across all age groups in one particular year. For a given year, it represents the average lifespan for a hypothetical group of people, if they experienced the same age-specific death rates throughout their whole lives as the age-specific death rates seen in that particular year.

Graph showing the rise in life expectancy around the world during and after the Gilded Age

Furthermore, life expectancy rose by over 50 percent in the following fifty years, while GDP really only seems to take off in the late 1900s, once the quaternary sector of society begins to take hold, further showing the missed encapsulation of the truly valuable increase in life expectancy (Dattani, 2023; Rockoff, 2008).

Overall, GDP is a great way to measure the economic growth concerned with the general quantity and amount of goods in transactions, but it is limited in its ability to capture the quality and value of these goods. Additionally, lifestyle changes are not included within these GDP metrics. Like the health and productivity portions of lifestyle mentioned earlier, the increased urbanization and also interconnectedness between the population prove another aspect of the limits of GDP. Society was completely transformed into a more connected community rather than large, separated farms. GDP completely ignores this transformation and instead focuses on the side effects, such as the new forms of housing, like apartments and bungalows.

The United States experienced some of the fastest growth in the world in the late 1800s and early 1900s, even compared to Europe. One of the main differences between the two regions was the adaptability to new inventions. While in the United States, new inventions were quickly adopted and used in daily life, these same inventions generally took longer to permeate throughout Europe. The United States also came ahead in many of the important inventions that revolutionized the world. For example, the electric lightbulb, crucial for the productivity of time, especially at night, was invented in the United States, quickly gaining widespread adoption and greatly helping the population. In Europe, its usage was impeded by the older infrastructure that was unable to support the widespread use of electricity (A short history of electric lighting - colite technologies, 2024).

The telephone, crucial for communication, was also invented in the United States, by Alexander Graham Bell, a Scottish-born inventor, illustrating the progress made in the US from foreign-born immigrants (Ericsson history: the invention of the telephone, 2025; 1870s – 1940s: telephone | imagining the internet | elon University, 2025). Due to the telegraph and government regulations, it was slow and fragmented to spread throughout Europe (Telephones | encyclopedia.com, 2025). In its earliest years, there were as many telephones in New York state as all of Europe combined, with the telephone being vital for business and personal use (Gordon, 2017).

On the other hand, the automobile was invented in Europe, and greatly changed the transportation landscape, but production in the US soon overtook Europe, along with its mass adoption and improvements (Eli et al., 2025). Henry Ford's Model T in the United States further revolutionized people's access to automobiles and allowed a large portion of the population to experience its effects (Eli et al., 2025). This was far from Europe, where cars remained mainly for the wealthy and elite (Eli et al., 2025).

These changes in the American economy completely revamped society as people knew it and lives were completely revolutionized. New commodities and conveniences invented changed the entire structure of where people lived, shifting from the previously common rural farm to the new urban cities with much more access to the new technology of the time. This shift to the

cities also allowed more people to experience the transformation of society at the same time, such as streetcars, and also access to public conveniences like hospitals and grocery stores, creating overall economic improvements largely ignored by GDP, and therefore the common conversations about the growth of the American economy.

Works Cited

1. Digital history, 2025. https://www.digitalhistory.uh.edu/era.cfm?eraID=9&smtid=1.

2. R. J. Gordon. The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War. Princeton University Press, 2017. https://doi.org/10.2307/j.ctvc77bwm.

3. S. Dattani, L. Rodés-Guirao, H. Ritchie, E. Ortiz-Ospina, M. Roser. Life expectancy. Our World in Data. 2023,.

4. U. Sinclair. The jungle. CreateSpace Independent Publishing Platform, North Charleston, South Carolina, 2021.

5. Colonial and pre-federal statistics, 2025.

6. World bank open data. World Bank Open Data https://data.worldbank.org, 2025.

7. Measuring worth - graphs. https://www.measuringworth.com/graphs/, 2025.

8. J. Rosenbloom. The colonial American economy, 2018.

9. P. C. Mancall, J. L. Rosenbloom, T. J. Weiss. NBER working paper series, 2025.

10. R. Shackleton. Total factor productivity growth in historical perspective, 2013.

11. H. Rockoff. Great Fortunes of the Gilded Age. pg. w14555, http://www.nber.org/papers/w14555.pdf 2008 https://doi.org/10.3386/w14555, 2008.

12. A short history of electric lighting - colite technologies. https://colitetech.com/blog/a-short-history-of-electric-lighting/ 2024.

13. Ericsson history: the invention of the telephone. ericsson.com https://www.ericsson.com/en/about-us/history/changing-the-world/phones-for-everyone/the-invention-of-the-telephone, 2025.

14. 1870s – 1940s: telephone | imagining the internet | elon university. https://www.elon.edu/u/imagining/time-capsule/150-years/back-1870-1940/, 2025.

15. Telephones | encyclopedia.com. https://www.encyclopedia.com/history/encyclopedias-almanacs-transcripts-and-maps/telephones, 2025.

16. S. Eli, J. K. Hausman, P. W. Rhode. The model t. The Journal of Economic History. Vol. 85, pg. 110–151, 2025, https://doi.org/10.1017/S0022050725000014.

Hopping Through the Hoops: A Comparative Study on the Socio-Cultural Impact of Basketball in North America, China, and Europe By Lucas Li

Abstract

This research paper will comparatively explore the nature of three different basketball environments and their impact on Europe, America, and China. Primarily, it will focus on player experiences, social, economic, cultural, institutional, and governmental features, and impacts. We will first discover the varying influence of the NBA and American basketball on the previous features mentioned, to analyse its current situation. How the previous US domination in basketball has globally affected basketball culture, while using significant examples like Allen Iverson to illustrate how American hip-hop and basketball culture have influenced current basketball game style and fashion trends. After that, we would focus on the cultural differences of each community and how regional culture relates to fame and money. To emphasise how a historical sports background and regional fashion trend impact the community. In saying so, we will consider the popularity of other cultures and sports: Football in Europe and the popularity of basketball within China. The regional culture, being closely related to institutional and government influence, would lead us to our next point about how the basketball ecosystem is contrasted and how basketball is structured, including analysis such as coaschin culture, levels of professional, and school leagues, with further explanation on the benefits and drawbacks of each region. Thus, it also includes how the government's attention would influence the popularity and funding of the sport. These, however, would all be closely referenced to the NBA: the origin and most popular basketball league across the globe. This means that all of our investigations would be closely related to the NBA's cultural influence and its popularity play style within different countries, we would also look into regional players that we have mentioned who are in the NBA. Finally, I conclude by commenting that despite cultural and institutional variations, global basketball still asirationally and financially revolves around the American ecosystem.

Introduction

Basketball, being one of the top 10 most popular sports around the world, has gained an incredible amount of growth in its popularity through the catastrophic growth of the NBA and regional adaptation to basketball around the World (SportsGeeks, 2025). The significant amount of growth in basketball popularity has a strong relationship with the growth of technology. This has helped basketball and the NBA spread virally around the World through social media. In this article, we will discuss basketball and the NBA's impact around the World through examples.

In doing so, we would use basketball in the United States, Europe, and China as regional and country examples to highlight the popularity and ways of popularity in these areas. Starting with the United States, basketball is one of the 'big four' sports in North America, alongside football, baseball, and ice hockey has a great influence and popularity within the state (Wilson, 2025). Especially because having the biggest basketball league in the World has gathered an intense and popular basketball culture in the United States while considering the United States

being one of the first Countries to play basketball. Basketball in Europe is slightly less popular due to the multi-sports culture in different regions and the domination of soccer. But basketball is significantly popular in various countries like Spain, Greece, Lithuania, etc. The government and population also take basketball as a serious sport and have a long history of playing basketball, being competitive in international events for a long time. Lastly, basketball is a massive sport in China, arguably the most popular sport after table tennis and badminton. It also has a huge NBA fanbase, and players like Yao Ming have elevated it massively, estimated to have over 300 million people playing basketball recreationally (Uptin Saiidi, 2018). While being a big, growing country, China has huge popularity and potential within sports, especially basketball. Therefore, I will comparatively analyse the growth of basketball in these regions and countries through 3 different analytical categories: culture and socio-economic factors, institutional influence, and affinity to the NBA and American basketball. This would help people understand the huge potential of the basketball industry and understand how culturally different these regions are while producing different types of pro players, and how governmental status would also affect the basketball performance of the countries. We would then expand on how culturally and institutionally basketball would influence the countries and how the country's specific culture and institutional status would affect basketball growth in each country.

Culture, being a definite crucial part affecting regional basketball, has led to significant variations among different players due to their regional culture and play style. As culture shapes values, beliefs, and behaviours, different cultures can influence what players enjoy and how to engage in playing basketball. In countries that value team sports or athleticism, basketball thrives in countries like China and the US. Where basketball is often fast-paced and showy in the US due to its heuristic features in its culture, but in regions like Europe and China, team basketball is emphasized due to its traditional culture and value on different play styles. In addition, regional culture also affects fan base and fan behaviours with different reactions. This would influence the player's play style based on the fans' encouragement for criticism. For example, if the fan base gives really extreme behaviours, players tend to be more excited and extreme. Culture also affects the basketball system and society through the role model's influence, spreading societal values to the basketball system. 2 Great examples are Iverson and Yao Ming, 2 superstars with significantly different personalities who have influenced their society differently and affected society in different ways (Lev et al., 2023).

Institutions are equally important to regional basketball development and influence the basketball system to a large extent. Institutions like government, legal systems, education, and sports bodies set the rules for how society functions. While also being a part of the cultural factor, these institutions influence how people act and the general personality of the country. They could enable or restrict opportunities: funding, development programs, access to courts, and coaches can heavily influence the regional basketball system. It also helps ensure fair play, safety, and organisation of the basketball culture. Thus, strong institutions like the NCAA in the US or national federations in Europe provide clear paths from youth to pro levels and affect basketball growth within the country (What Is the NCAA College Basketball Academy? |

NCAA.com, 2023). Where governments or private institutions invest in sports facilities and programs to different extents, it helps basketball development in these regions, while different decisions lead to significantly different basketball systems. The set rules also shape how the game is played, enforced, and promoted internationally. Different rules from the FIBA and the NBA have led to 2 different play styles with significantly different rules to the game (L. G, 2024). Where institutions like the NBA create global outreach programs, like basketball without borders, helping to spread basketball to places like China, Africa, and Europe, has illustrated the influence of institutions. In saying so, the NBA has had a huge impact on current basketball play style and culture due to its popularity worldwide. Players like Allen Iverson, Kobe Bryant, and Stephen Curry have significantly changed the way basketball is played.

In summary, the global rise of basketball is not simply a story of athletic excellence but a complex interplay of culture, institutions, and the international influence of the NBA. By exploring basketball's development in the United States, Europe, and China, we can understand how deeply culture shapes the play style, fan behaviour, and societal impact of the sport in each region. At the same time, institutions — from powerful leagues like the NBA to local governments and development programs — play a crucial role in structuring access, growth, and the professionalism of basketball worldwide. The NBA's global outreach and the role of iconic players have accelerated the game's popularity, influencing local cultures and inspiring millions of new players. Through this comparative analysis, we will see how the cultural identity of each region, combined with the strength and strategy of its institutions, has contributed to basketball's explosive growth and regional diversity. Ultimately, this will highlight not only the global potential of the basketball industry but also the unique ways in which different societies have adopted and adapted the game, producing distinct basketball cultures and pathways to success.

Institutional and Governmental Influence.

Institutions such as systems and structures that organise sports, including schools, professional leagues, government agencies, and training programs, are a crucial part when it comes to basketball. The institutions build the country's basketball system, which is a big part of the country's basketball development. As they create specific programs and policies for players to adapt to, they create an environment that the majority of local players will need to adapt to. Therefore, it could hugely impact regional basketball development. To be specific, the US, Europe, and China have different institutions, where the USA has institutions like the NCAA, AAU, and the NBA to organise games for players within the institutions. Whereas Europe and China mainly rely on club fixtures, having institutions like FFBB, Euroleague, and CBA. This could cause different advantages and challenges to these institutions. For example, institutions in the USA normally have high exposure and publicity to players who are in the leagues and have created fully commercialised institutions like OTE and more, which could bring more publicity to the players and put the spotlight on these players. Europe has a long-term development program that would efficiently build basketball superstars like Luka Doncic, who is from Real Madrid's basketball club; however, the low amount of exposure might lead to many great players

not having the ability to play in the highest league and earn enough money for a living. Similarly, China has a fully structured basketball pipeline with early scouting and centralized training, where the over-centralisation could limit the players' freedom and harm their passion towards basketball from a small age. This illustrates how crucial institutions are for national basketball development.

Firstly, diving into American basketball institutions, they are formed by many commercialised leagues that organise players to play for teams that compete for a trophy. Their youth basketball program starts in institutions like the NCAA and the school system. Where there is a smooth transition from high school to college, and to the highest league being the NBA, where their ability to continuously produce the best performing players has shown that America has a mature basketball system. Similarly, there are also institutions like the AAU (Amateur Athletic Union), which is similar to other youth programs but is a privatised youth league with intense scouting and promotes early exposure to players. However, there are concerns about whether early exposure to youth basketball players can cause burnout and distract their passion for basketball, which could be considered a drawback of the system. With the ultimate goal for all players being that the NBA is located in the USA, it being a central dominant league across the world has a highly structured system with significantly strong financial power, which makes it the "highest level of basketball" in global basketball development (NBA Teams Surpassed US$ 11.3 Billion in Revenue in 2024; Total Valuation Reached US$ 132.8 Billion – Sports Value, 2024). The NBA has a highly commercialised system which built many sports celebrities like Michael Jordan, Kobe Bryant, and LeBron James, which brings a large amount of financial profit to the players within the league and provides huge exposure to these players being the best basketball players on Earth. Despite all these advantages, there are concerns towards the NBA that the act of over-commericalising the league will slowly lose competitiveness of games and how teams and players will act in the market's favor to play in certain styles. This could be significantly harmful while considering the significant amount of criticisms the audience of the NBA has launched on the league, and causing a decrease in its attraction towards basketball players.

Then, focusing on the European basketball institutions are mainly covered by clubs, public funding, and sports academies, and also talking about how school and sports are separated. The European clubs system is significant in countries like Spain, France, and Lithuania, where players join clubs early and move through development tiers and finally into the highest level (EB Advisory Group, 2025; Grokipedia, 2025). The club system clearly focuses on long-term development and fundamental skills of each player with minimal exposure (Basketball Vibes, 2025). Some significant players who are from the European club system are Luka Doncic and Paul Gasol. While public funding and sports academies also build up major parts of European basketball institutions in many countries where they have government-supported youth programs, which made it more accessible and lowered the economic barrier for entering sports programs (EB Advisory Group, 2025; Basketball Embassy, 2020). They often share more inclusivity and a more systematic talent development program

since it is funded by governments and is managed professionally. Building on national federations and professional leagues, the national teams program, and how federations manage player development through selecting highly professional basketball players in regional teams as representatives in European competitions. An example would be the EuroLeague, where it is pan-European and less commercial than the NBA, while still remaining a high level of competitiveness to select a present target for regional European players to aim towards (EuroLeague, n.d.). Another great feature of the European basketball institutions is how school and sports are separated, unlike in the US. School-based sport is not dominant, where most players join institutions that we have mentioned before, like clubs and regional government-funded institutions, instead of attending school sports programs — a model frequently seen in Sweden and other European countries (The Basketball Embassy, 2020). This changes how players balance academics and sport, and clearly separates the two to provide more structured timings for players to organise their personal academics and sporting careers. Therefore, this highlights the key features of European basketball institutions and how they are mainly divided into a club system and a government-funded system, different from the U.S.

Now, coming to the features of the Chinese basketball institutions and their impact on Chinese basketball development. The Chinese basketball institutions are mainly divided into state controlled sports systems, CBA, schools and national training camps, and partnerships with the NBA. A major player in the Chinese basketball system is the government and state-controlled sports system, where the government invests in training centers, facilities, and athlete scouting. They mainly focus on the basketball development of the entire state or country, and focus on elite development and national pride instead of helping youth development. Where the highest performing Chinese basketballers often play in the CBA (Chinese Basketball Association), which is controlled by the government, and Yao Ming's reforms. However, it being the highest level of basketball in China still lacks the commercial polish and competitiveness when compared to the NBA and Euroleague. Despite the presence of top-level and pro players, many Chinese institutions also focus on youth development programs such as school training, national training camps, and partnerships with the NBA (Weinstein, 2014). These programs tend to develop the most talented players and are less academically focused, with success defined by sporting achievement. Where school leagues are a major part of Chinese youth basketball development, it has a reasonable amount of commercial potential and spotlights on the players, which, to some extent, is an advantage of the Chinese youth development program. Another path of youth development is the partnerships with the NBA. The NBA has invested in grassroots training like Basketball Without Borders and other programs to develop great players who do not have the opportunity to perform in the CBA, and has shown a hybrid approach of government infrastructure and American branding (Weinstein, 2014). This illustrates that the basketball institutions in China are mainly top-down controlled and limit player autonomy; however, they also have a decent amount of exposure and a great youth development program for their future in basketball.

Now, comparing the features of the different institutions, discussing which system is better or worse in different areas. Despite all systems being very inclusive and including a major portion of the population attending basketball development programs. By numbers, the Chinese basketball institutions include basketball fans and players of 600 million and 300 million, respectively (Sprung, 2019). Only 31 million people in the US have played basketball at least once in their lifetime, which is larger than the European basketball involvement population (Broughton, 2022). Therefore, the Chinese basketball system is more inclusive; however, a limitation that could be mentioned is the large amount of total population that China has compared to the other regions. Although China has a more inclusive basketball system, the United States still remains the top leader for producing the most NBA talent globally. This could be mainly due to the local influence that the NBA has and the history and experience in producing NBA players. However, this could soon be challenged by the significantly increasing European NBA players in the league. 3 of the 5 in the top MVP list in 2025 are non-American, while having 2 European players (Kadlick, 2025). This illustrates how European institutions are catching up in the ability to produce high-quality NBA talent at the highest level of basketball globally. While comparing government institutions and private-led systems, both sharing significant benefits, the Government led system, mainly in China and Europe, tends to be more inclusive, especially at the youth levels, and also providing public sports schools and state funding can help players from all socioeconomic backgrounds. Like in France, public sports academies like INSEP develop talent regardless of income. Whereas private-led systems, mainly in the US, offer elite opportunities, but access often depends on money, connections, or location, meaning that it is harder to access the pro basketball system in the US. Players in lower-income communities may struggle to afford travel teams (AAU), camps, or gear. Therefore, success often hinges on getting noticed and has a high-risk, high-reward model. In saying so, all three types of basketball systems and institutions have their own style and benefits for local basketball development.

Institutions play a vital role in shaping basketball development across the US, Europe, and China, each with distinct systems reflecting their cultural, economic, and political contexts. The US relies on highly commercialised, private-led institutions like the NCAA, AAU, and NBA, offering significant exposure and financial rewards but raising concerns about player burnout and competitive balance. Europe's club-based, government-supported system prioritizes long-term development, inclusivity, and a clear separation between academics and sport, producing technically skilled players, though with less media spotlight and commercial appeal. China's state-controlled model focuses on elite national development, with centralized training and government funding, providing broad access due to its large population but potentially limiting individual freedom and creativity. While China leads in sheer participation numbers, the US still dominates in producing top NBA talent, though Europe is rapidly closing the gap with its structured clubs and academies. Government-led systems tend to be more inclusive and equitable, whereas private-led models in the US offer elite but less accessible pathways. Ultimately, no system is perfect; each has unique strengths and challenges that shape player

development, exposure, and basketball culture. Understanding these institutional differences is crucial for appreciating how regional basketball ecosystems evolve and compete on the global stage.

Socio-Economic Factors

Now moving on to the socio-economic factors for evolving in basketball. Basketball, compared to other more luxurious sports like golf, is quite a significantly cost-effective sport since all that is required are shoes, a ball, and a rim. However, the setup cost for a professional basketball court is still high. Equipment needed to play basketball includes a good basketball ($20-$50), appropriate footwear ($120-$300), and an appropriate basketball court ($8-$15 per time), but these costs vary due to different locations and the development of the areas. Costs like transportation costs for basketballs and shoes may vary from other locations. Places with low manufacturing fees, like China, may enjoy a lower price for basketballs and shoes, whereas regions like Europe and the US may have a higher price in order to purchase these products. This is, however, different for courts and places to play basketball. This is because as the development of the country grows, the maturity of the country's sports facilities will be more widespread, and the community may enjoy better facilities with more social welfare (Emergen Research, 2025). In this case, China, compared to Europe and the US, may have fewer widespread basketball facilities for society, and the amount of government-owned sports facilities may also be lower due to the amount of development. But what does this mean? In response to this, a higher barrier to access basketball due to its higher difficulty of living or higher price to play basketball, may influence low-income families and individuals' ability to access and play basketball.

Where financial benefits are crucial to individuals and every citizen, the ability to gain financial benefits from playing basketball, and the costs of playing basketball are significant for people. Therefore, many countries have taken sports seriously as a career and provided many work opportunities in sports in order to encourage more players while encouraging economic growth. Using the NBA as the ultimate example of financial success. Players in the NBA are receiving a large amount of salary and income from both club contracts and even sponsorships: A top-tier NBA player can earn an average salary of over $10 million USD, with superstars earning far more (Mustansar, 2025). This can also be seen in many different regions, like Australia, China, and Europe, where they all have professional leagues like the NBL, CBA, and the Euroleague… The growth of these global leagues has created a middle-class for professional basketball players, even if they don't reach the NBA, talented players can still earn a comfortable living in their home country or overseas. Thus, off-court earnings also hold a significant amount of income from a player's salary, where endorsements and media deals also exist in these other leagues and bring economic benefits for players. This suggests that nowadays, people could have the dream of playing basketball for a living and making it their career.

Furthermore, the perception of basketball has evolved into the "cool" sport category, and it is now associated with a wide range of social groups and how accessible everyone it accessible to join. Initially, basketball was often a middle-class activity played in urban gyms, and its early

reputation was linked to its invention by a Canadian physical education instructor, James Naismith. However, as basketball became more and more popular while raising star players, the perception evolved with the rise of famous players like Michael Jordan, whose global stardom transcended sport and made basketball a symbol of high-end consumerism and athletic excellence (Arash F, 2025). Contrast this with the sport's deep roots in urban culture and its association with a more diverse range of demographics. The quick growth in popularity of basketball has soon influenced urban culture and has become big in many developed regions like Europe, China, and the US, where the game style, music, and fashion have been heavily influenced by urban and hip-hop culture (mainly rooted in the US) (Toxigon, 2025). This has made the sport a global cultural phenomenon and has a large social impact in developed regions. Therefore, basketball has successfully demonstrated how it has evolved from a simple gym activity to a globally recognised symbol of style, talent, and cultural diversity.

As basketball can be played in both everyday and luxurious environments, for example, street courts and exclusive elite facilities, it demonstrates how it can exist in both public and private, financially exclusive settings. Furthermore, basketball, as a generally simple game, needs only a ball and a hoop, making it incredibly accessible. Public street courts and community centres have made the game a communal pursuit, which has made it many countries' first choice in building sports facilities compared to high-cost sports like golf and football. As developing or developed countries, this becomes a great tool in fostering a sense of community and teamwork regardless of players' economic background. In addition, basketball has also become a path to "change your life," as we've seen too many stories of people with poor economic and social status, who, through playing basketball, have become some of the wealthiest people in the world. This suggests how basketball offers a tangible path to a better life, where pathways like scholarships at high school and university are great examples of how it could directly change a person's life. Therefore, this has made basketball a popular sport in many regions due to its special features, where the three regions both share common factors of needing the opportunity or light for people to see a unique pathway in their life.

The media has been a defining force in shaping basketball's culture and transforming it from a North American pastime into a global phenomenon. The rise of television broadcasting in the 1980s and 1990s, especially through ESPN and CCTV in China, allowed international audiences to witness the sport's athleticism, creativity, and emotional intensity. Documentaries such as The Last Dance (2020) later redefined basketball narratives, portraying it not merely as a sport but as a story of ambition, struggle, and triumph. This emotional storytelling extended basketball's influence beyond the court, turning players into global icons and symbols of resilience. The digital revolution has since amplified this process: social media platforms like Instagram, TikTok, and YouTube have democratized basketball's reach, with stars such as LeBron James and Stephen Curry shaping public discourse through personal storytelling, activism, and lifestyle branding. In China, social platforms like Weibo and Douyin replicate this phenomenon by localizing global basketball narratives—fans follow NBA stars while also celebrating domestic heroes like Yao Ming and Zhou Qi. Thus, media representation has evolved

from showcasing athletic excellence to creating a shared global narrative that merges American individualism with localized interpretations of aspiration and identity.

Fandom and fashion further demonstrate basketball's cultural hybridity. In the United States, fandom often revolves around individual players, reflecting a celebrity-driven culture where loyalty shifts with athletes rather than teams. This "super-fan" culture, rooted in admiration for personal success and charisma, contrasts sharply with Europe's club-based tradition, where fans remain loyal to institutions such as Real Madrid or CSKA Moscow. In China, basketball fandom blends both models: while NBA players enjoy celebrity-level devotion, national and community pride remain central, seen in organized fan groups and patriotic support for the Chinese Basketball Association (CBA). This synthesis highlights how global fandom adapts to local values—American-style celebrity culture intertwines with Chinese collectivism, forming a hybrid model of cultural expression. Fashion, too, mirrors this global fusion. The rise of sneaker culture, pioneered by Nike's Air Jordan line, transformed basketball shoes into symbols of identity, aspiration, and style. In the U.S., sneakers represent empowerment and urban self-expression; in Europe, they have been absorbed into luxury streetwear through collaborations between sportswear brands and high-fashion designers. In China, local brands such as Li-Ning and Anta reinterpret basketball fashion by fusing Western street style with Chinese cultural elements and by signing global ambassadors like Dwyane Wade or Klay Thompson. Similarly, basketball jerseys and streetwear aesthetics have become global fashion staples, with local adaptations that reflect national pride and social belonging. Across all regions, basketball culture—through media, fandom, and fashion—functions as a stage for hybrid identities, where global symbols of sport are continuously reimagined through local creativity.

Conclusion

Ultimately, basketball has become a cultural hybrid as it evolved from North America and entered other regions like China and Europe. However, depending on different regions' original rooted culture and differentiated sporting character, basketball culture has evolved differently in America, China, and Europe. This research paper has investigated how institutional structures have shaped the development of basketball in these places, while America has professional school leagues like the NCAA and AAU. China and Europe have different structures; Europe has well-organised basketball leagues like the Euroleague and has built European club models based on that, whereas China is mainly based on a government-led system and promotes mass participation and national pride but restricts player autonomy through centralisation. This research paper has also considered the socio-economic factors' influence on basketball culture, as it influences accessibility and opportunity to play basketball. Although basketball is quite a resource-efficient sport, it still requires land and a certain amount of resources from organisations. This has caused governments with a limited amount of resources or a different sporting direction to limit citizens' access to basketball; for example, many European countries may lean on football as their "main sport," which leads to an unequal

distribution of football and basketball, which consequently leads to limited access to basketball. However, this research paper does contain some drawbacks due to the data scope and the limited and inconsistent amount of data in order to support a strong argument. In addition, the cultural generalisation may also be a factor that this paper has neglected by generalising places into America, Europe, and China, while specific regions of these catorgarised places could have slight differences in culture. While this paper acknowledges its limitations in data scope and cultural generalisation, the comparative framework remains valid because it draws upon widely recognised institutional models (such as the NBA, Euroleague, and CBA) and observable transnational trends in basketball's globalisation, allowing for meaningful generalisations that capture the broader cultural and structural dynamics shaping the sport across regions. This can lead to further discussion questions, like how does football influence the local culture of different regions?

Works Cited

Andrew, A., & Morris, M. (n.d.). "I 'I Believe You Can Fly': Fly": Basketball Basketball Culture Culture in Postsocialist Postsocialist China China. Retrieved May 13, 2025, from https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1030&context=hist_fac

Arash F. (2025, August). Michael Jordan Marketing Strategy (2025 Update): Jordan Brand, Nike & Global Influence. Brand Vision Insights. https://www.brandvm.com/post/michael-jordan-basketball-marketing?utm.com

Aslam, S. (2025, April 19). How European Coaches Develop Elite Basketball Players: A Guide. Basketball Vibes. https://basketballvibes.com/european-basketball-player-development/?utm.com

Bale, J. (2002). Sports geography. Routledge.

Boyd, T. (2003). Young, Black, rich, and famous : the rise of the NBA, the hip hop invasion, and the transformation of American culture. Doubleday.

Broughton, D. (2022, September 26). Hoops slam dunk: Basketball adds to long streak as favorite major sport to play. Sports Business Journal. https://www.sportsbusinessjournal.com/Journal/Issues/2022/09/26/Portfolio/Sports-particaipation/?utm.com

Cunningham, C. (n.d.). BASKETBALL BEDLAM IN BARCELONA.

Emergen Research. (2025, October 3). Basketball Gear Market In-Depth Growth & Competitive Strategy Report [2024–2034]. Emergen Research. https://www.emergenresearch.com/industry-report/basketball-gear-market?utm.com

ESPN. (2025). ESPN.com. https://www.espn.com/mens-college-basketball

Euroleague Basketball About - EuroLeague. (2023, March 15). EuroLeague. https://www.euroleague.com/euroleague-basketball/about/?utm.com

FIBA.basketball. (n.d.). FIBA.basketball. https://www.fiba.basketball/europe

Grokipedia. (1970, January 21). European professional club basketball system. Grokipedia. https://grokipedia.com/page/European_professional_club_basketball_system?utm.com

Guest, S. (2013, June 17). The Globalization Of The NBA. Forbes. https://www.forbes.com/sites/sap/2013/06/17/the-globalization-of-the-nba/?utm.com

How popular is basketball in france. (2016). Gdybl.org. https://gdybl.org/misc/how-popular-is-basketball-in-france.html

in. (2020, April 22). The Basketball Embassy. The Basketball Embassy. https://www.thebasketballembassy.org/posts/2020/4/22/transatlantic-youth-development-differences-in-approaches-between-sweden-and-the-us?utm.com

Ji, H. (2023). Research on the Economic Impact of the NBA Business League. Advances in Economics, Management and Political Sciences, 42, 148–151. https://doi.org/10.54254/2754-1169/42/20232099

Kadlick, M. (2025, May 22). Complete 2025 NBA MVP Voting Results: SGA Edges Out Nikola Jokic to Capture Award. SI; Sports Illustrated.

https://www.si.com/nba/complete-2025-nba-mvp-voting-results-sga-edges-out-nikola-jokic-to-capture-award?utm.com

Karamürsel, C. (2025). The Cost of Youth Development in European Basketball A detailed analysis of structural and financial challenges of developing talent in Europe. https://eb-advisorygroup.org/wp-content/uploads/2025/08/EBAG-Report_18_Cost-of-Youth-Development-in-European-Basketball_FINAL.pdf?ut.com

Kim, K. (2023). The Proceeding of 2023 4th Asia Sport Science Conference.

Krasnoff, L. S. (2017, December 26). How the NBA went global. The Washington Post. https://www.washingtonpost.com/news/made-by-history/wp/2017/12/26/how-the-nba-went-global/?utm.com

L. G. (2024, April 23). Unraveling the Court Discrepancy: FIBA vs. NBA Rules in Basketball. The World of Basketb. https://www.theworldofbasketball.com/post/unraveling-the-court-discrepancy-fiba-vs-nba-rules-in-basketball?utm.com

Lev, A., Orr Levental, & Tamir, I. (2023). From legends to legacy: the impact of fan influence on retiring athletes in premier league basketball. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1295487

Luo, L., Dai, Y., & Huang, F. (2015). Glocalization and the Rise of the Chinese Basketball Market. The International Journal of the History of Sport, 32(10), 1321–1335. https://doi.org/10.1080/09523367.2015.1066992

MIES (Switzerland) - A landmark research study about the state of basketball has confirmed the global interest in the sport and the rising attraction of the FIBA Basketball World Cup. (2024, June 28). Fiba.basketball. https://about.fiba.basketball/en/news/popularity-of-basketball-soaring-globally-highlighted-by-increased-interest-in-the-fiba-world-cup?utmhatgpt.com

Minoso, G. (2024, December 20). The Globalization of Professional Basketball: Context and Competition Matters in the NBA, WNBA, and Olympics. The Sport Journal. https://thesportjournal.org/article/the-globalization-of-professional-basketball-context-and-competition-matters-in-the-nba-wnba-and-olympics/?utm.com

Mustansar. (2025, June 24). NBA Salary in 2025 - Salary Solver. Salary Solver. https://salarysolver.com/nba-salary/?utm.com

NBA China and the Globalization of Basketball. (2016). Samford University. https://www.samford.edu/sports-analytics/fans/2016/nba-china-and-the-globalization-of-basketball?utm.com

NBA Teams Surpassed US$ 11.3 Billion in Revenue in 2024; Total Valuation Reached US$ 132.8 Billion – Sports Value. (2024). Sportsvalue.com.br. https://www.sportsvalue.com.br/en/nba-teams-surpassed-us-11-3-billion-in-revenue-in-2024-total-valuation-reached-us-132-8-billion/?utm.com

SportsGeeks. (2025, October 13). Top 10 Most Popular Sports in the World 2025. SportsGeeks. https://sportsgeeks.net/most-popular-sports-in-the-world/?utm.com

Sprung, S. (2019, March 4). Inside The NBA's Push To Make Basketball The World's Most
Popular Sport. Forbes.
https://www.forbes.com/sites/shlomosprung/2019/03/04/nba-china-ceo-derek-chang-take
s-us-inside-nbas-push-to-make-basketball-worlds-most-popular-sport/?utm.com

Toxigon. (2025, March 22). Basketball and Its Impact on Urban Culture. Toxigon.
https://toxigon.com/basketball-and-its-impact-on-urban-culture?utm.com

Turner, T. (2014). German Sports Shoes, Basketball, and Hip Hop: The Consumption and
Cultural Significance of the adidas "Superstar", 1966–1988. Sport in History, 35(1),
127–155. https://doi.org/10.1080/17460263.2014.931293

Uptin Saiidi. (2018, November 20). The NBA is China's most popular sports league. Here's how
it happened. CNBC.
https://www.cnbc.com/2018/11/20/the-nba-is-chinas-most-popular-sports-league-heres-h
ow-it-happened.html?utm.com

Weinstein, B. (2014, October 17). NBA, Chinese Ministry of Education announce partnership to
grow basketball in China - NBA.com: NBA Communications. NBA.com: NBA
Communications. https://pr.nba.com/nba-china-ministry-education-partnership/?utm.com

What is the NCAA College Basketball Academy? | NCAA.com. (2023). NCAA.com.
https://www.ncaa.com/collegebasketballacademy/about?utm.com

Wilson, D. (2025, May 13). CEOWORLD magazine. CEOWORLD Magazine.
https://ceoworld.biz/2025/05/13/most-popular-sports-in-the-united-states-2025/?utm.com

The Notorious Web of Deception
Unmasking the World's Most Infamous Financial Scams By Neev D. Karia

ABSTRACT

In this paper, I will consider five major contemporary financial scams and provide reasons why the perpetrators were able to orchestrate these crimes despite a relatively strict level of scrutiny and the rule of law. These major financial scams have been committed in different countries and locations. The 1st scam that I will be taking an insight into is the Charles Ponzi scam, which was committed in the United States of America. This scam is famously known as the 'Ponzi Scheme'. Next, the Harshad Mehta Scam, which was committed in India. Harshad Mehta was a small-town boy who wanted to become a stockbroker, and the story goes on like this. Following this scam was the famous Jordan Belfort's penny stock scam, which was committed in the USA. Following that, we have the Lehman Brothers scam, which was also a scam committed in the USA, and it is also the most ruthless banking scam in history, as it became a major reason for the 2008 crisis. Lastly, we have the Truong My Lan scam, which was committed in Vietnam. This scam might be unfamiliar, but it is known to be the biggest scam in Southeast Asia. I will begin my analysis by first providing a brief historical overview of how these scams unfolded. As opposed to other widespread cases of corruption and collusions that have come out of these locations, these scams primarily capitalized on hidden loopholes within the political and economic systems of the host countries. These individuals, with the help of corrupted government officials and/or officers, were able to take advantage of these loopholes in the governmental systems. I will finally end the paper by commenting on the long-lasting effects of these scams on financial markets and the economy, both at the national and global levels.

INTRODUCTION

In a world where every single penny counts, financial scams are invisible traps that shatter the dreams of individuals and families and ruin their lives (1). Financial scams are usually defined as a deliberate attempt to trick an individual or organization into giving money or personal information, usually for personal gain. Financial scams can be defined as modern-day traps where individuals or organizations manipulate information to create illusions of opportunities, and following that, illegally stealing assets is sometimes more valuable than an individual's whole life. These scams are not only a type of crime but also negatively affect the victims mentally and physically. It makes the victims lose loads of money, sometimes their entire life's savings, and also affects them emotionally through consequences such as anxiety and

[1]

---

[1] **1.** Investopedia. "The Collapse of Lehman Brothers."
11. Dalal & Basu, *ThePrint*.
7. MoneyShow. "How the Belfort Scam Worked."
24. National Archives.
25. OpenAI. ChatGPT Conversation with Neev Karia.

depression. In this paper, I will highlight the modus operandi of five major scams, namely the Ponzi Scam, Jordan Belfort Scam, Lehman Brothers Scam, Truong Lang Scam, and lastly the Harshad Mehta Scam. In doing so, I will also underscore why these perpetrators were able to commit these crimes with relative impunity for a while, and finally end the paper with my comments on the normative detestability of financial scams due to the harm they cause to the respective economies.

In the early 1920s, the scam by Charles Ponzi occurred, and this scam is famously known as the 'Ponzi Scheme'. It was not just a scam but a trend setter for similar financial delinquencies to come. Charles Ponzi was a person who kept things simple. In the early 1900's the international postal reply coupons were introduced to the world. These IRCs acted as vouchers that could be exchanged for postage in other countries. For example, if you send a letter to a friend in a foreign country, and do not want them to pay for the postal services for sending a reply letter back to you, then you put this voucher in the mail that you send, and then they can redeem the voucher from their country's post office. Here, Charles Ponzi told several investors about this investment opportunity and how they could earn money by buying these vouchers from countries that have a lower rate and selling the vouchers in countries that have a higher rate. Like this, Charles got his initial investors and promised them to double their money. Now, here is where the term Ponzi scheme was invented. Ponzi actually never invested the money into the IRCs; he actually used the money and paid back the investors by taking money from other investors. With the money he received, he lived lavishly and fulfilled his dreams. He was later caught because there were no investors who were ready to invest in his so-called "investment opportunity". Next, the 1992 Harshad Mehta scam. Harshad Mehta, aka the Big Bull, was a stockbroker who found a loophole in the Indian banking and stock market system. He bought the bank managers and printed fake bank receipts and diverted those bank funds into the stock market. During that time in India, banks borrowed money from other Indian Banks; they had to do this via a stockbroker and had to compulsorily issue BRs (Bank Receipts). Now the catch was that the BRs stayed with the stockbroker for about 2 weeks before the money was transferred to another bank. During these 2 weeks, Harshad Mehta used to divert this money to the stock market and artificially inflate the price, and then make money out of it. [2]

Following that is the greatest known scam ever, the Jordan Belfort Scam, which is famously known as the Penny Stock scam. This scam was exposed by the release of the movie "The Wolf of Wall Street". Jordan was a perfect salesman, and using this ability, he was able to start and run a brokerage firm called Stratton Oakmont. Jordan, with the help of his salesmen employees, engaged in artificially inflating the stock prices and then selling the stocks before they crashed, leading to defrauding investors of about $200 million. Belfort was later convicted of securities fraud and money laundering and served 22 months in prison. Next is the Lehman Brothers Collapse. The Lehman Brothers invested heavily in mortgage-backed securities, and this was one of their main reasons for their collapse. Mortgage-backed securities are a big pool

[2] **9.**"1992 Indian Stock Market Scam." *Wikipedia*.
11. Dalal, Sucheta, and Debashis Basu. "The Harshad Mehta Story: How It Unfolded."
12."Scams That Changed India's Capital Market." ResearchGate.

of loans bundled together and then sold to investors. The Lehman Brothers heavily invested in subprime mortgages (they are high-risk loans given to people with poor credit scores) and also sold these subprime mortgages to investors. Whereas this was not the scam, the scam was when the Lehman Brothers hid their true financial conditions using a trick called 'Repo 105'. This accounting trick allowed the Lehman Brothers to make their finances look much healthier than they were, and this misled investors, regulators, and credit rating agencies. Last is the Saigon Commercial Bank (SCB) scam by a Vietnamese woman called The Truong My Lan, and it is known as Vietnam's biggest financial scam. Truong was not part of the SCB group, but she was alleged to control it secretly. She controls the SCB through proxies and manipulates its lending operations. She was also alleged to have orchestrated fake loans, which helped her withdraw a vast amount of funds from the bank. (Will need to find how to bring up a connection from this)

While these scams were being orchestrated, it is also important to look at the perspective of the country's financial situation during that time. Charles Ponzi was an Italian immigrant who moved to the United States of America. His scam in 1920 brought up various consequences and changes to the State's financial system, especially in terms of regulatory changes, public trust, and the awareness of financial fraud. Being an immigrant, the fraud-fueled nativism and anti-immigrant sentiments in some circles, and people became more aware and alert about such financial fraud. In the short term, thousands of ordinary Americans lost their precious life-saving, and there were a few bank failures because such banks were highly exposed to Ponzi schemes. It brought about a ripple effect on various local economies. In short, the overall United States at that time was in bad off.

Next, we move on to explore another scam, one that was orchestrated by someone infamous in the Indian financial markets as the Big Bull. After the scam occurred and he was exposed, the Indian stock market crashed, and retail investors suffered huge losses. Apart from that, several public sector banks like SBI had illegally transferred money to Mehta, and such banks suffered huge losses(like SBI lost over ₹600 crore). There was a huge banking system crisis in India. The scam forced Indian markets to modernize and made the banking laws tighter. Following that is the Jordan Belfort scam. [3]His scam exposed how loosely regulated the counter markets were, and the USA government brought in stricter laws and rules to the markets. The SEC (The Securities and Exchange Commission) closely started monitoring penny stock trading in particular. The stock market crashed to some extent, and many small-cap market indexes, like the NASDAQ. The SEC also made broker licensing and monitoring stronger. Overall, the US economy was damaged, but the government and the SEC also became more proactive on such cases. Next is Vietnam's largest financial fraud, orchestrated by Trương Mỹ Lan between 2012 and 2022. The case, being one of the largest financial frauds in Vietnam, affected Vietnam's private banking system by a lot. This was because people lost trust in private banks, and they shifted to state-owned banks.

---

[3] **5.**Investopedia. "Who Is Jordan Belfort?"
6."Jordan Belfort." Wikipedia.
7.MoneyShow. "How the Belfort Scam Worked."
3.Sorkin, Andrew Ross. "Anatomy of a Meltdown." The New Yorker.
1.Investopedia. "The Collapse of Lehman Brothers."

This scam slowed and shrank Foreign Direct Investment (FDI) by a lot, too. Lastly, it is the Lehman Brothers Collapse — a collapse that did not just shake the US but also the whole world. When the Lehman Brothers filed for bankruptcy, there was a worldwide panic that led stock markets to crash all over the world, and the banking system collapsed. Thousands of Lehman Brothers' employees became unemployed, and the global unemployment rate spiked because the collapse had caused severe damage to other sectors, such as real estate and manufacturing. Lastly, this collapse was one of the main causes of the Great Recession of 2008. This collapse helped the US to make stricter laws on banks that allowed the government and customers of the bank to a greater transparency into the bank that they are giving their money to.

FINANCIAL SCAMS: A SAGA OF MANIPULATING FINANCIAL MARKETS AROUND THE WORLD

These three scams — the Harshad Mehta scam, the Jordan Belfort scam, and the Lehman Brothers collapse — are grouped because they all took advantage of weaknesses in financial markets and securities. Even though they happened in different countries and involved different tricks, they all show how stock markets and financial systems can be misused to make huge illegal profits. Unlike normal bank frauds or sales scams, these cases were focused on manipulating stocks, fake investments, and using financial loopholes. This grouping helps us see how important it is for governments to watch over stock markets strictly. It also shows us that when people in power cheat, it can damage entire economies, hurt millions of small investors, and destroy trust in the financial system. These scams prove that financial crimes are not only about stealing money but also about destroying people's dreams and affecting their mental health through stress, anxiety, and depression.

Harshad Mehta's scam in India is often called the "Big Bull scam." During the early 1990s, India's financial system was still developing, and there were many loopholes. Mehta, who was a stockbroker, used fake bank receipts (BRs) to get huge amounts of money from banks. He convinced bank officials to give him these receipts and then diverted that money into the stock market. He used the borrowed money to buy large amounts of shares in certain big companies, which caused their stock prices to go up sharply. [4]
When normal people saw these prices going up, they thought the market was strong and safe, so they also started buying these shares. Once Mehta sold his shares at high prices, the prices crashed, and many people lost their life savings. This showed how one person's greed could damage an entire financial system and destroy the lives of thousands of small investors.

The Jordan Belfort scam in the United States is famously known as the "penny stock scam." Belfort was a very talented salesman who started a brokerage firm called Stratton Oakmont. Together with his employees, he tricked people into buying cheap, worthless stocks by

---

[4] **16.** France 24. "Vietnam's Multi-Billion-Dollar Fraud Case: Key Things to Know."
15. Al Jazeera. "Vietnam Court Upholds Tycoon's Death Sentence…"
13. "Vạn Thịnh Phát Fraud Case." Wikipedia.
14. "Trương Mỹ Lan." Wikipedia.
4. The Brookings Institution. "History Credits Lehman Brothers' Collapse…"

making them seem like great investment opportunities. This method is called a "pump and dump" scheme. They first "pumped" up the stock prices by convincing people to buy, and when the price was high enough, they "dumped" their own shares and made huge profits. After they sold, the stock prices crashed, and normal investors were left with heavy losses. In total, Belfort's scam defrauded investors of about $200 million. He used this money to live an extremely luxurious life with expensive cars, private jets, big parties, and drugs. Belfort was later caught and served about 22 months in prison for securities fraud and money laundering. His story became even more famous after the movie The Wolf of Wall Street was released.

The Lehman Brothers' collapse in 2008 is one of the biggest financial failures in history. Lehman Brothers was a large investment bank in the United States. The bank invested heavily in mortgage-backed securities, which are bundles of home loans sold to investors. Many of these loans were "subprime," meaning they were given to people with poor credit scores and were very risky. Lehman Brothers kept buying and selling these securities and made a lot of money at first. However, to hide their real financial problems, they used a trick called "Repo 105." This trick made their balance sheet look stronger than it actually was and fooled investors and regulators. When people started defaulting on their home loans, the value of mortgage-backed securities dropped quickly. Lehman Brothers could not handle the losses, and in September 2008, it declared bankruptcy. This triggered a huge financial crisis all over the world. Many people lost their jobs, homes, and savings.

The people behind these scams had different personalities but similar greed. Harshad Mehta was born in 1954 in Gujarat, India, and moved to Mumbai to work as a stockbroker. He was known for his charm and was seen as a market genius for some time. However, his entire image was built on cheating the banking system and fooling normal investors. Jordan Belfort was born in 1962 in New York and started his career selling meat and seafood door-to-door. Later, he used his selling skills to build Stratton Oakmont. He loved to show off his wealth and partied nonstop. Richard Fuld Jr., the CEO of Lehman Brothers, was born in 1946 and worked at Lehman for many years. He was known for being tough and extremely ambitious. His push for bigger and riskier profits without listening to warnings led to Lehman's collapse.

These men all used the money they stole or gained illegally to live extremely lavish lives.[5]Harshad Mehta bought many expensive cars, luxury apartments, and continued investing more to keep his scam going and make himself look like a market hero. Jordan Belfort spent his money on crazy parties, drugs, yachts, designer clothes, and fancy houses. He wanted to live like a king and didn't care about the people who lost everything. At Lehman Brothers, the money was not used for just one person's personal luxury but was reinvested to grow the business in an aggressive and risky way. Executives received huge bonuses and paid themselves big salaries, while normal investors and employees suffered in the end.

---

[5] **4.**The Brookings Institution.
1. Investopedia. "The Collapse of Lehman Brothers.
7.MoneyWeek. "The Rise and Fall of Jordan Belfort."
11. Dalal & Basu. *ThePrint*.
10."Harshad Mehta." *Wikipedia*.

When we look at these scams together, it is clear that unchecked greed and weak regulations can destroy not only individual lives but also entire economies. The negative impact of these scams shows how a lack of greed controls the complete downfall of economies. The global recession was created by the downfall of Lehman Brothers, while Harshad Mehta's manipulation of the stock market and Trương Mỹ Lan's banking fraud were damaging to India and Vietnam, respectively. The financial public trust these events lost was devastating, and the financial savings lost were enough to force the government into tougher regulations.

These stories teach us that financial markets must always be strictly monitored and that people should be careful before trusting someone who promises quick and easy profits. At the end of the day, these scams show us that ethics and honesty are much more valuable than any amount of money.

All three of them were eventually caught, but all of them were caught in very interesting ways. Firstly, Harshad Mehta. He was a hero for all the stock brokers and bulls in the stock market, except one journalist, Sucheta Dalal. She was always shocked about Harshad Mehta's super luxurious life, which was common for a stockbroker during those days, but she knew that something was being cooked backstage. She started investigating on her own and started digging deeper, and then found out that he was using fake bank receipts to get money from the banks and push up stock prices. She released a news article exposing everything, and then the banks started checking the accounts and the ledger books, and found out that they had been really cheated. Now the authorities had no choice but to conduct various income tax raids and arrest him. Following that is the activity and the firm that Jordan Belfort ran, parallel drew attention to him across financial networks. His firm, Stratton Oakmont, came into notoriety when there were too many complaints about scams filed by investors. With time, the U.S. SEC and the FINRA started investigating Belfort for scam operations involving penny stocks as well as fraud. Ultimately, Belfort's business partner decided to reveal details that led to his arrest. Thorough investigations revealed that he had indeed indulged in scams, fraud, and money laundering. Contrary to Jan Belfort, financial systems work differently for larger firms. The monopoly that Lehman Brothers had also attracted attention, which resulted in it being more covered than others. In 2008, their set investments in mortgage resources began crashing, along with the housing market. [6]

This eventually lost Lehman the trust of investors, banks, and financial institutions permanently, exposing their deep-rooted issues to the financial world. This exposed the debt of their balance sheet that they strategically covered. Lehman was ultimately forced to file for bankruptcy and ceased all functionalities. They were later studied on how they used "Repo 105" to manipulate balance sheets and deceitfully disguise much of their debts to seem healthy.

Now it is very important to know why these financial scams were orchestrated by the individuals. Firstly, Harshad Mehta, who was known as the Big Bull, wanted to become rich very quickly and wanted a big fame and name very quickly. The loopholes that he had found in the

---

[6] **11.**Dalal, Sucheta. ThePrint.
5.FINRA / SEC references (via Investopedia "Jordan Belfort").
3.Sorkin, Andrew Ross. The New Yorker.
1.Investopedia. "The Collapse of Lehman Brothers."

banking system made him have an easy income that motivated him to orchestrate the scam, as he wanted to get rich quickly. Following that is Jordan Belfort, who was always motivated by extreme greed and a luxurious, part-filled lifestyle. He knew that he had excellent selling skills, which he used to cheat investors and earn a ton of money. Lastly, he, too, like Harshad Mehta, wanted quick money, and he thought that using his excellent salesman skills, he would be able to earn very quickly and a lot of money, which he eventually did, but in the wrong way. Lastly, Richard Fuld and the top executives at Lehman Brothers wanted to keep up with other big banks and show huge profits every quarter to impress shareholders and earn big bonuses. They ignored long-term risks and believed they were "too big to fail," taking reckless bets, thinking someone would always rescue them if things went wrong.

In summary, Harshad Mehta, Jordan Belfort, and The Lehman Brothers all show us what happens when greed and the desire for quick success take over honesty and responsibility. They each wanted to become rich and famous as fast as possible, even if it meant breaking laws and ruining the lives of innocent people. By cheating the financial systems, they hurt small investors, destroyed trust in stock markets, and caused huge damage to their countries' economies. Their stories teach us an important lesson that no matter how tempting it may look, choosing money over ethics and harming others is never worth it in the end.

NON-FINANCIAL SCAMS: A SAGA OF GREED, INFLUENCE, AND COLLAPSED INSTITUTIONS

These 2 scams — The Truong My Lan Banking Scam and the Charles Ponzi Scam — are clubbed together because of their similarity in taking advantage of non-financial markets and securities. Both these scams use tricks that they had up their sleeves and played with the stock market. These scams are separated from the others due to their simplicity of orchestration. Unlike the other 3 frauds in the section above, who exploited the stock market and investment banks, Ponzi and Lan did not operate through the financial market instruments like securities. These scams were not technical manipulations but rather based on system deception and exploiting the trust of individuals and organizations. This clubbing helps us to identify how simple frauds are sometimes not considered, and it's too late when they are discovered. This clubbing benefits the analysis by showing that financial fraud does not always require access to the market — sometimes all it takes is public trust and unchecked authority. [7]

To review how Charles Ponzi orchestrated the scam, we need time travel to the 1920s. In the 1920s, a lot of regulations and rules were not present that are present in the world today. These regulations will be discussed in the upcoming paragraph, but a main point to note is that the U.S Securities and Exchange Commission (SEC) did not exist until 1934. This means that during the time of Charles Ponzi, there was no dedicated authority that monitored investment schemes, making the fraud happen much more easily. As no authority looked at the activities the

[7] **10.** "Harshad Mehta." Wikipedia.
6. "Jordan Belfort." Wikipedia.
7. MoneyWeek. "Rise and Fall of Jordan Belfort."
1. Investopedia. "The Collapse of Lehman Brothers."

investors did, Charles used a very simple formula to earn money, as we talked earlier. Giving you a recap that Ponzi promised people a 50% profit in 45 days using International Reply Coupons. He told the investors that he would buy the IRCs in countries with weak currencies and sell them in the USA for profit. This is where the plot twist comes in. He really never invested the money into IRCs, and not only IRCs, but anywhere else, too. It is important to note that people wanted quick money after WWII, and that is, they were falling into the pockets of players such as Charles Ponzi. Coming back to his simple formula, he took money from the investors and used that money for his personal needs and wants. When the time came to pay the investor with his money, he took investment from other investors and repaid it, creating an illusion of high returns. In his prime, Charles was making about 250 bands a day.

Charles Ponzi's scheme was simplistic, and he found a loophole, whereas Truong My Lan had to surreptitiously break several established laws. Lan controlled the Saigon Joint Stock Commercial Bank (SCB) from behind the scenes despite not being on the board officially. By bribing regulators, auditors, and a government official who received $5 million to hide bad debts, she was able to approve illegal loans and funnel them into her real estate empire known as Van Thinh Phat Holdings Group. It is very important to know how big the scandal really was. Before the scandal broke, SCB was considered one of the largest commercial banks in Vietnam. In fact, in mid-2022, SCB ranked fifth among Vietnam's largest banks in terms of total assets. Following that, she approved over 2500 illegal loans, which totaled around $44 billion between the years 2012 and 2022. Her operation always used to work in a tightly connected circle, which involved fake companies, family members, and corrupt insiders. Think of this scam as the layers of the Earth. The scam was so deeply embedded in the internal structure of the bank that it did not trigger any red flags on SCB, and regulators who would look at SCB's operations would think that everything was going fine due to the proper paperwork and approved loan.

Moving on, we will look into a little bit of the biography of these fraudsters. Born in Italy in 1882, Charles Ponzi came from a poor family and struggled with low-paying jobs and a few cons that made him little money before launching the major Ponzi scheme. In 1903, he migrated to America to pursue the American Dream. He was always known for his charm, confidence, and the ability to convince people and win people's trust despite his criminal background. Due to his personality of getting really quick, he became a symbol of greed and deception in 1920s America. [8]

He arrived in Boston with only a few dollars and worked odd jobs, including as a dishwasher and bank clerk. As days passed by and his release of the Pozi scheme made him around $15 million, which is worth around $200 million in today's world. Next is Truong My Lan, who was born in 1956 in a Chinese-Vietnamese business family in Ho Chi Minh City (Saigon). She began her career in the 1990s by selling cosmetics and property. She later cofounded the Van Thinh Phat Group, one of the largest real estate empires in Vietnam, in the

---

[8] **20.**"Charles Ponzi." Wikipedia.
22. Smithsonian Magazine. "In Ponzi We Trust."
14."Trương Mỹ Lan." Wikipedia.
17. VnExpress. "All Central Bank Inspectors Took Bribes from SCB."

early 1990s. Over the years, she built her powerful and low-profile figure in the Vietnam business world with ties across several industries. She was considered untouchable due to her wealth, donations, power, and political influence. She was later arrested and still serves her sentence in prison.

The jobs that they have done were quite risky, and something that they wanted to get out of it was money. With the help of the money, both of them wanted to complete their lavish wants. Firstly, Charles Ponzi, who always wanted to enjoy the American dream lavishly, was due to the things that he had gone through in his childhood. Ponzi never wanted to buy one thing in particular with his money – instead, he wanted to live like a wealthy and respected man and gain social recognition. After earning money, he was considered a successful businessman, but he too bought a few things. A huge mansion in Lexington, Massachusetts, expensive custom-made suits and fine jewelry, and a chauffeur-driven limousine. He also hosted lavish parties and gatherings at his mansion, using these events to build trust, showcase his success, and capture more people's attention. With the money, he also tries to buy a bank and a shipping line to appear more legitimate. Overall, Ponzi used his money very wisely sometimes, but sometimes he threw it on anything he liked. In contrast, Truong My Lan, unlike Ponzi, was extremely private and low-profile and did not throw frequent public parties or flaunt her wealth in a flashy way. Her way of spending money was quite different from Ponzi's. She funneled her wealth that she earned into luxury real estate, high-end hotels, and expensive properties, often hidden behind layers of shell companies. This means that the properties were taken in the name of different shell companies. Both had different approaches to spending their wealth in terms of power and luxury.

We know how well they were able to spend the money that they stole, but it is very important to know how their scams were exposed or how they were caught. Ponzi and Truong were both caught in very different ways, and this shows how differently they operated. Ponzi's idea was a short-term plan and was based on a basic idea that was simply known to never last forever. He promised huge profits through international reply coupons, but he could not keep up with the big group of investors. Once it dried up, red flags rose, and that caused journalists and investigators to find out that the volume of IRCs did not match Ponzi's claims. The Boston Post published an expose which caused an investigation to run in his company. There was an official audit against his company, and this put him behind bars. On the other hand, Truong was far more complex. Her fraud of quietly taking out billions of dollars as loans had remained undetected for several years. She was later caught when an investigation was run on SCB. Regulators noticed that SCB was issuing a large number of loans to companies that were either fake, closely linked to each other, or had no real operations. [9]

It was also noticed that SCB had grown quickly after merging with other banks, but the balance sheets did not show any healthy operations that were taking place.

[9] **22.**Smithsonian Magazine.
23. Biography.com.
17. VnExpress.
16. France 24.
18. TheInvestor.vn.

All of these suspicions led to an investigation in which Truong was caught. While Ponzi's scam blew up in public, Lan's was slowly uncovered after several years, showing how different types of fraud require different ways to get detected.

Both Cahles and Truong faced serious punishments for their crimes. Ponzi was arrested in 1920 with multiple mail fraud charges. He was sentenced to 5 years of federal prison and ordered to pay about $150 in court fees. After his federal sentence, he was prosecuted by Massachusetts on larceny (theft of personal property) charges. He was then sentenced to 7-9 years of state prison in Charlestown. Ponzi was never able to repay his investors – he had stolen $15 million (which is over $200 million today), but was broke when caught. After his release from prison, he was immediately deported to Italy in 1934. He later moved to Brazil, where he lived in poverty until he died in 1949. He may be dead, but his name lives on, in fact, the synonym for fraudulent investment scheme – "Ponzi Scheme". On the other hand, Truong My Lan's case was far more recent and much larger than Ponzi's. In 2024, she was sentenced to death by the Ho Chi Minh City Court, found guilty of playing with $12 billion through $44 billion in bad loans. It was not only she who was convicted; 80 others, including her husband and top SCB executives, were also convicted. She is currently in prison, and there is international attention set on her, on whether her sentence will be carried out or reduced.

Both Charles Ponzi and Truong My Lan were able to carry out the scams because, for both of them, one thing was understood very well: people are willing to trust when they see the promise of big rewards. Ponzi played with the investors' greed and optimism, convincing people who wanted to make quick money that they had found a golden opportunity to double their money in a short period of time. On the other hand, Lan was able to use her institutional power, political connections, and influence over the banking system to carry out the scam. Weak regulatory systems and poor oversight over the systems made these frauds plausible. In Ponzi's time, the US had no laws regulating securities and investment schemes, which allowed him to operate only for several months before being challenged. There are several laws and key safeguards that did not exist in 1920. For example, there was no US Securities and Exchange Commission (SEC) until 1934, so there was no dedicated authority to monitor investment schemes. Ordinary people have very little financial knowledge, and there were no public warnings about fraud. In those times, there were no emails or real-time auditing that would help spot the inconsistencies — the records were maintained in a ledger or a book. There was no social media or TV, so news did not travel so easily, meaning that scams could not be exposed quickly. It was very hard for banks to communicate effectively with each other, making it hard to verify Ponzi claims of holding international reply coupons. It was like everything was missing from the world that we are currently in, including Anti Money Laundering (AML) and Know Your Customer (KYC) laws. Post World War I, the economic climate created depression for quick riches. [10]

[10] **23.** Biography.com. "Charles Ponzi."
24. National Archives. "'When Ponzi's Bubble Burst.'"
14. "Trương Mỹ Lan." Wikipedia.
16. France 24.

Now think about this: a person willing to give the last money you own without taking interest. Isn't it intriguing? What Ponzi was able to do was make this fraud work. He took advantage of the depression led by World War I and created his wealth.

Both were also talented or skilled at building an image that built trust. Ponzi had the image of being a successful immigrant – a person who is lavishly living the American Dream –, wearing fine clothes, jewelry, and throwing parties made them believe that he was a financial genius and knew something that they did not. Lan had a quiet and elegant public profile as a respected businesswoman, using forged documents and complex banking operations to hide her crimes from the public. Lan was also able to orchestrate such a scam due to the corruption in Vietnam. She bribed several people including a Vietnamese government official. The key bribery recipient was Đỗ Thị Nhàn, who served as the head of Banking Inspection and Supervision Department II under the State Bank of Vietnam. He was allegedly bribed by Lan's steam with $5.2 million — sometimes in hidden boxes disguised as fruit – to manipulate bank inspection reports and suppress revelations of SCB's misuse of funds. Both of their confidence, salesmanship, and deep understanding of legal and financial loopholes allowed them to fool investors, employees, and authorities to allow them to orchestrate the scam. In both cases, the truth was kept hidden and was not discovered as the victims were blinded by the power and success that they possessed. All the victims were into the "fear of missing out" as Ponzi promised impossible profits, while Lan used her knowledge to thrive in the real estate boom in Vietnam, making her projects "once in a lifetime" investments. Together, their stories of orchestrating such frauds show how unchecked ambition, combined with weak checks and balances, can lead to disasters that could damage the entire economy.

In summary, the scams orchestrated by Charles Ponzi and Tuong My Lan were, although separated by a century and continent, share the same roots of unchecked greed and systemic failure. Ponzi targeted people who wanted to make quick money and lured them with promised profits through a nonexistent arbitrage scheme. While Lan bribed and manipulated the Vietnamese banking system to funnel billions of dollars for her real estate empire. One was dependent on public enthusiasm and hype, whereas the other was based on the corrupted systems. Yet both showcase a common ground of showing how individuals can abuse trust. Ponzi was exposed when Math did not Math and Lan fell due to the determined corruption probe. In short, their legacies remain a cautionary tale for the world, talking about the human cost of financial deception and the importance of vigilant regulation.

CONCLUSION

This paper has explored the 5 most notorious financial scandals: the collapse of Lehman Brothers, the Jordan Belfort stock market scam, the Harshad Mehta securities scam, the Truong My Lan banking scam, and the Charles Ponzi investment scheme. These all occurred at different times and in different parts of the world. [11]But there is at least one similarity: all of them were the

---

[11] **24.**National Archives.
22. Smithsonian Magazine.
15. Al Jazeera.

product of unmitigated greed, lack of thorough and effective regulation, and misplaced public trust in all the individuals and institutions that were sufficiently dominant and apparently very successful. Consulting all of them in conjunction shows just how financial systems can be abused and exploited when there is unmitigated ambition and a disengaged regulatory environment that fails to keep track of fast-paced developments.

The first three cases—Lehman Brothers, Jordan Belfort, and Harshad Mehta—were grouped due to their close relations with financial markets and securities. The collapse of Lehman Brothers in 2008, due to years of reckless mortgage-backed securities investments, was the trigger of the global financial crisis. Jordan Belfort's firm, Stratton Oakmont, was a front for stock price manipulation and employed fraudulent sales tactics to defraud unknowing investors. Harshad Mehta, on the other hand, abused and exploited the loopholes in India's banking system to create a fake and temporary prosperity by inflating stock prices. All three of them used the financial market. Greed and weak institutional control were similarly manipulated by them both, regardless of the hundred-year difference. Emulating Ponzi fraudulently profiting off of international reply coupons and preying on the weak institutional controls of the United States financial system. Emulating Ponzi fraudulently profiting off of international reply coupons and preying on the weak institutional controls of the United States financial system. In the 21st century, Lan would politically and financially instigate the embezzlement of 12 billion dollars from the Saigon Commercial Bank. Ponzi and Lan had embezzled 12 billion dollars from the Saigon Commercial Bank. Controlling the fucking 12 billion dollars from the Saigon Commercial Bank.

While comprehensive, this paper inevitably has some limitations. Given the great differences in contexts, some of the five major scandals to be examined simply could not be covered in complete depth. Some technical financial details had to be summarized, including the derivatives that Lehman Brothers used and the layered fraudulent loans that Lan used in her schemes. The study is also based predominantly on secondary sources such as reports, court documents, and reputable news sources, which may be limited by partial perspectives or under confidentiality restrictions. Moreover, each country's social and political contexts were.

This paper may have some limitations, but it also has meaningful strengths for readers today. By placing five different scams side by side, it helps the audience recognize more universal patterns of financial deception. The author identifies overconfidence, inadequate regulations, and public greed as examples of factors that potentially fit each case, helping the readers avoid isolating each case. Additionally, the author consolidates the historical and contemporary contexts by illustrating how the same behaviors persist despite the evolution of financial systems. These readers, especially young investors, students, and policymakers, can assist in the active prevention of future scams by recognizing the early warning signs of highly promised returns and institutional opacity, as discussed.

---

19. Saigon Giai Phong.

This paper teaches us that, regardless of age or location, financial scams will flourish where ambition and opportunity intersect, and where oversight is absent. Thankfully, it also shows us that learning from the past is, and always will be, a strong defense. By recounting the five scams, the paper has, in great part, helped the audience become more aware of, more critical of, and more responsible about the intricacies of the financial world. This is far more than just a recounting of five infamous financial scandals. It serves as a warning against the repetition of history and a call to remain vigilant.[12]

---

[12] **1.**Investopedia. "The Collapse of Lehman Brothers."
11. Dalal & Basu, ThePrint.
7.MoneyShow. "How the Belfort Scam Worked."
24. National Archives.
25. OpenAI. ChatGPT Conversation with Neev Karia.

Works Cited

1. Investopedia. "The Collapse of Lehman Brothers." Investopedia, IAC Publishing, 17 Sept.2008,https://www.investopedia.com/articles/economics/09/lehman-brothers-collapse.asp. Accessed 10 July 2025.

2. "Bankruptcy of Lehman Brothers." Encyclopedia Britannica, Encyclopedia Britannica, Inc., https://www.britannica.com/event/bankruptcy-of-Lehman-Brothers. Accessed 12 June 2025.

3. Sorkin, Andrew Ross. "Anatomy of a Meltdown." The New Yorker, 1 Dec. 2008, https://www.newyorker.com/magazine/2008/12/01/anatomy-of-a-meltdown. Accessed 5 Sept. 2024.

4. The Brookings Institution. "History Credits Lehman Brothers' Collapse for the 2008 Financial Crisis — Here's Why That Narrative Is Wrong." Brookings, https://www.brookings.edu/articles/history-credits-lehman-brothers-collapse-for-the-2008-financial-crisis-heres-why-that-narrative-is-wrong/. Accessed 20 Oct. 2025.

5. Investopedia. "Who Is Jordan Belfort?" Investopedia, IAC Publishing, https://www.investopedia.com/investing/who-is-jordan-belfort/. Accessed 25 Sept. 2025.

6. "Jordan Belfort." Wikipedia, The Free Encyclopedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/Jordan_Belfort. Accessed 11 May 2025.

7. MoneyShow. "How the Belfort Scam Worked." MoneyShow, 22 Feb. 2023, https://www.moneyshow.com/articles/tradingidea-60444/. Accessed 2 Oct. 2024.

8. Kranish, Michael, and Joe Keohane. "The Rise and Fall of Jordan Belfort." MoneyWeek, https://moneyweek.com/512249/great-frauds-in-history-jordan-belfort-and-stratton-oakmont. Accessed 19 Aug. 2025.

9. "1992 Indian Stock Market Scam." Wikipedia, The Free Encyclopedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/1992_Indian_stock_market_scam. Accessed 28 July 2025.

10. "Harshad Mehta." Wikipedia, The Free Encyclopedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/Harshad_Mehta. Accessed 9 June 2024.

11. Dalal, Sucheta, and Debashis Basu. "The Harshad Mehta Story: How It Unfolded." ThePrint,2022,https://theprint.in/theprint-profile/scapegoat-or-mastermind-of-1992-scam-harshad-mehtas-fall-from-grace-still-haunts-family/1059017/. Accessed 15 Apr. 2025.

12. "Scams That Changed India's Capital Market." ResearchGate, 2023, https://www.researchgate.net/publication/373552466_Scams_That_Changed_India%27s_Capital_Market. Accessed 23 Sept. 2025.

13. "Vạn Thịnh Phát Fraud Case." Wikipedia, The Free Encyclopedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/V%E1%BA%A1n_Th%E1%BB%8Bnh_Ph%C3%A1t_fraud_case. Accessed 17 June 2025.

14. "Trương Mỹ Lan." Wikipedia, The Free Encyclopedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/Tr%C6%B0%C6%A1ng_M%E1%BB%B9_Lan. Accessed 1 July 2024.

15. Al Jazeera. "Vietnam Court Upholds Tycoon's Death Sentence for US $12 Billion Bank Fraud."AlJazeera,3Dec.2024,https://www.aljazeera.com/news/2024/12/03/vietnam-court-upholds-tycoons-death-sentence-for-12bn-bank-fraud. Accessed 9 Oct. 2025.
16. France 24. "Vietnam's Multi-Billion Dollar Fraud Case: Key Things to Know." France 24, 12Apr.2024,https://www.france24.com/en/live-news/20240412-vietnam-s-multi-billion-dollar-fraud-case-key-things-to-know. Accessed 7 Aug. 2025.
17. VnExpress. "All Central Bank Inspectors Took Bribes from SCB: Police." VnExpress International,https://e.vnexpress.net/news/business/companies/all-central-bank-inspectors-took-bribes-from-scb-police-4679807.html. Accessed 30 May 2025.
18. TheInvestor.vn. "Death Sentence Upheld for Vạn Thịnh Phát Chairwoman Trương Mỹ Lan."The Investor,https://theinvestor.vn/death-sentence-upheld-for-van-thinh-phat-chairwoman-truong-my-lan-d13617.html. Accessed 2 Mar. 2025.
19. Saigon Giai Phong. "Court Upholds Death Sentence for Trương Mỹ Lan." SGGP Online, https://en.sggp.org.vn/court-upholds-death-sentence-for-truong-my-lan-post114233.html. Accessed 18 Oct. 2025.
20. "Charles Ponzi." Wikipedia, The Free Encyclopedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/Charles_Ponzi. Accessed 14 Oct. 2025.
21. U.S. Securities and Exchange Commission. "Ponzi Schemes." Investor.gov, U.S. SEC, https://www.investor.gov/introduction-investing/investing-basics/glossary/ponzi-schemes. Accessed 3 Apr. 2025.
22. Smithsonian Magazine. "In Ponzi We Trust." Smithsonian Magazine, https://www.smithsonianmag.com/history/in-ponzi-we-trust-64016168/. Accessed 9 June 2024.
23. Biography.com. "Charles Ponzi – Scheme, Movie & Wife – Biography." A&E Networks, https://www.biography.com/crime/charles-ponzi. Accessed 21 Aug. 2025.
24. National Archives. "'When Ponzi's Bubble Burst.'" Prologue Magazine, National Archives,https://www.archives.gov/publications/prologue/2010/summer/ponzi-inmate-case-file. Accessed 12 Sept. 2025.
25. OpenAI. ChatGPT Conversation with Neev Karia about Global Financial Scams and Analysis. ChatGPT, OpenAI, 20 Apr. 2024–21 Oct. 2025, https://chat.openai.com/

How Different Factors Affect the Performance of High School Students
By Isbah Almas Maknojia

Abstract

The objective of this study was to examine how various factors affect the performance, productivity, and health of high school students through a series of self-reported weekly polls. The study was centered around understanding students' health, stress levels, and how other related factors may affect their academic and emotional stability throughout the semester. Data were collected through an initial poll about demographics followed by weekly surveys, which were organized and analyzed in a spreadsheet in order to study qualitative and quantitative results. The participants ranged from grades 9-12, ages 14-18, with diverse cultural and academic backgrounds and states of residence.

The results show that students who sleep more hours per night typically have fewer stressors, a lower stress level, a higher average mood, and self-report an overall state of well-being, with the opposite being true for students who sleep fewer hours per night. Sleep is essential for development during adolescence, as shown and supported by the data.

Future research may be conducted on how students' well-being may improve with a lighter load of schoolwork. Overall, the study shows that students experience higher rates of well-being if presented with more sleep and fewer stressors.

Keywords sleep, students, health, academic performance, stress, productivity

Introduction

My own experience inspired this research project. As an Honors and AP high school student, sleep and health can often become neglected due to external factors, such as stress from school and extracurricular activities. This topic is important because lack of sleep can cause detrimental changes in performance, physical health, and mental health. Since teenagers are under immense pressure due to school and their future, their health may not come first, which can be extremely degrading because they need a minimum of 8 to 10 hours of sleep per night. Only a small percent of total students may fulfill this requirement, which is extremely important to their health. The purpose of this research project is to find how different factors affect performance, mood, health, and more related to everyday life in high school students.

Hypothesis: Younger students will get more sleep each night and therefore, have an overall better mood, performance, health, etc. Students who sleep more on average will have an overall better mood, performance, health, etc.

Methodology

There were a total of 60 participants from grades 9-12, varying in age, sex, and race. Google Forms were utilized to collect data from each participant throughout the semester. There were two types of forms: an initial poll and a weekly poll. The initial poll was to be filled out

before the weekly polls were set to begin. Weekly polls were set to be filled out every Friday, as a reflection of the week. Quantitative and qualitative responses were collected. Quantitative responses that included a range were converted into a numeric. All data was organized into a spreadsheet based on the academic 6 weeks calendar. All data was converted using formulas and analyzed using pivot tables and charts. Any personal information linked to participants will not be included in the report, due to promising confidentiality from the public.

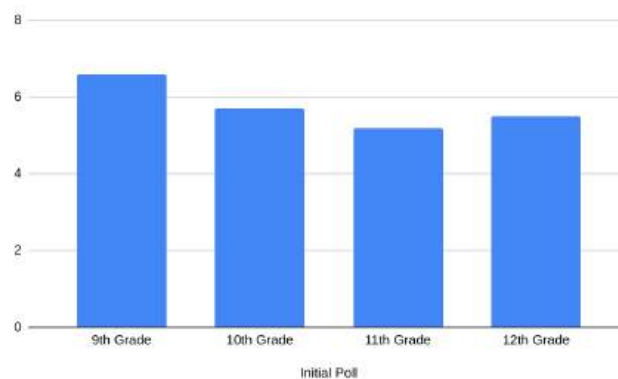## Quantitative Results

Initial Poll Results

Fig 1: Numeric of how many hours of sleep each grade typically gets per school night
The average amount of hours of sleep on school nights for 9th graders was 6.6, for 10th graders was 5.7, for 11th graders was 5.2, and for 12th graders was 5.5. On average, 9th graders get the most sleep every school night. On average, 11th graders get the least amount of sleep every school night.
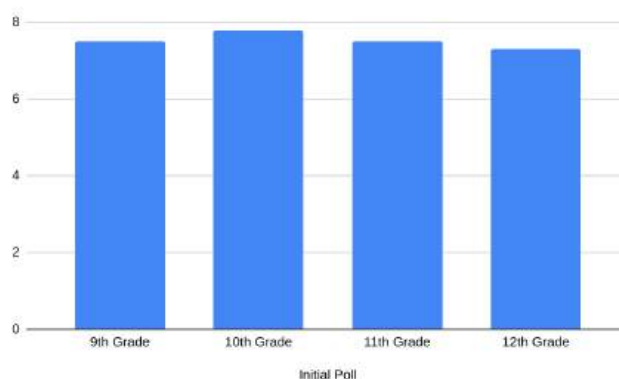

Fig 2: Numeric of how many hours of sleep each grade gets every weekend night
The average amount of hours of sleep every weekend night for 9th graders was 7.5. For 10th graders, it was 7.8 hours. For 11th graders, it was 7.5 hours. For 12th graders, it was 7.3 hours. On average, 10th graders get the most sleep every weekend night. On average, 12th graders get

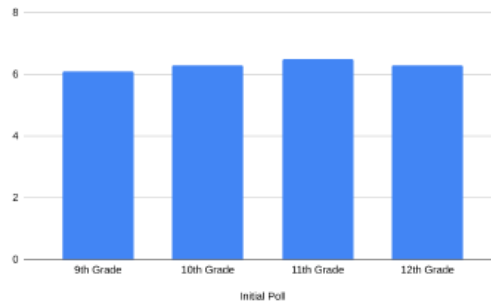the least amount of sleep every weekend night.

Fig 3: Rating of quality of sleep across grades

The average rating of quality of sleep for 9th graders was 6.1, for 10th graders was 6.3, for 11th graders was 6.5, and for 12th graders was 6.3. On average, 9th graders have the worst quality of sleep. On average, 11th graders have the best quality of sleep. A lower amount of sleep is associated with a better quality of sleep, while a higher amount of sleep is associated with a worse quality of sleep.
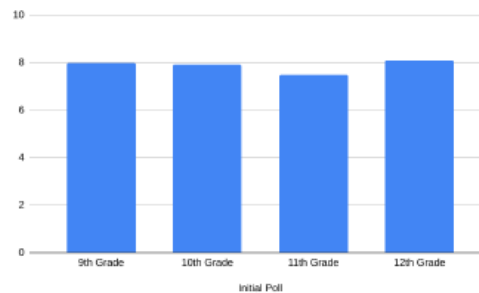
Fig 4: Rating of academic performance during Semester 1 across grades

The average rating of academic performance for Semester 1 for 9th graders was 8, for 10th graders was 7.9, for 11th graders was 7.5, and for 12th graders was 8.1. On average, 11th graders had the worst rating of academic performance. On average, 9th graders had the best rating of academic performance. More sleep is associated with a higher academic performance rating. Less sleep is associated with a lower academic performance rating.

Fig 5: Numeric of how many hours per week each grade spends on school outside of the classroom

The average number of hours per week spent on school outside of the classroom for 9th grade was 3.9, for 10th grade was 5.3, for 11th grade was 6.8, and for 12th grade was 6.3. On average, 11th graders spend the most time on school outside of the classroom. On average, 9th graders spend the least amount of time on school outside of the classroom. More sleep is associated with less time spent on school outside of the classroom,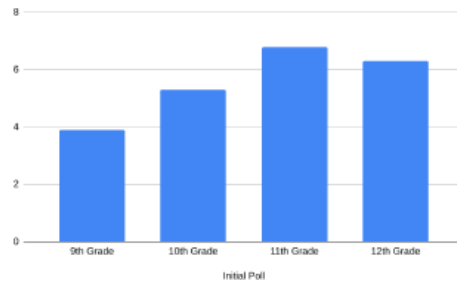 as well as a higher academic performance rating. Less sleep is associated with more time spent on school outside of the classroom, as well as a lower academic performance rating.
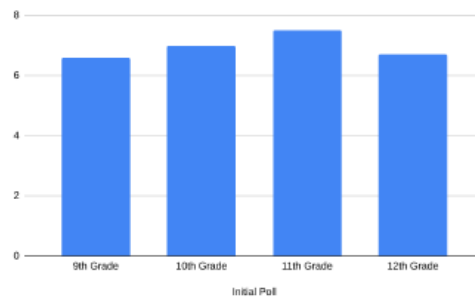
Fig 6: Rating of typical stress levels across grades

The average stress rating level for 9th grade was 6.6, for 10th grade was 7.0, for 11th grade was 7.5, and for 12th grade was 6.7. On average, 11th grade had the highest average stress level. On average, 9th grade had the lowest average stress level. A higher stress level is associated with less sleep and more time spent on school outside of the classroom. A lower stress level is associated with more sleep and less time spent on school outside of the classroom.
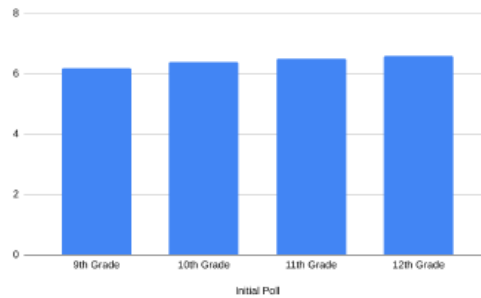
Fig 7: Rating of typical mood across grades

The average typical mood rating for 9th grade was 6.2, for 10th grade was 6.4, for 11th grade was 6.5, and for 12th grade was 6.6. On average, 9th grade had the lowest typical mood rating. On average, 12th grade had the highest typical mood rating.

Weekly Poll Results
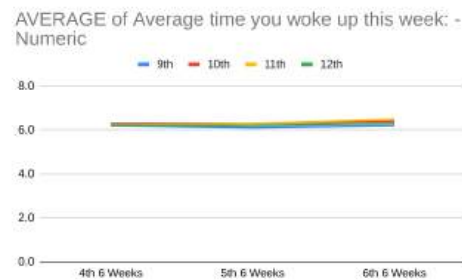
Fig 8: Numeric of average of average time students from grades 9-12 woke up throughout each week per semester
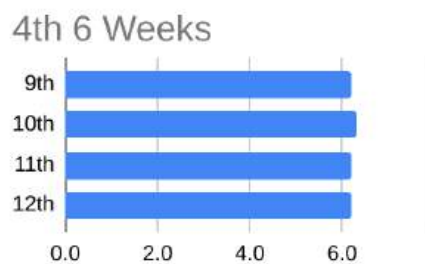
Fig 9: Numeric of average of average time students from grades 9-12 woke up throughout each week during the 4th 6 weeks
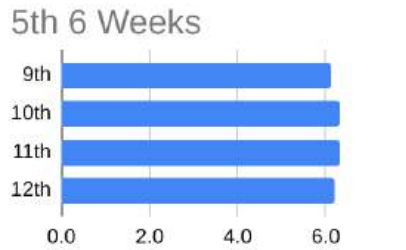
5th 6 Weeks

Fig 10: Numeric of average of average time students from grades 9-12 woke up throughout each week during the 5th 6 weeks
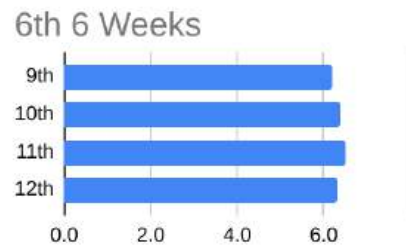
6th 6 Weeks

Fig 11: Numeric of average of average time students from grades 9-12 woke up throughout each week during the 4th 6 weeks

From the 4th 6 weeks to the 6th 6 weeks, the average time students across all grades woke up each week stayed consistent, with a few minor fluctuations based on grade. For 9th grade, it was 6.2 during the 4th 6 weeks, 6.1 during the 5th 6 weeks, and 6.2 during the 6th 6 weeks. 9th grade experienced a minor fluctuation during the 5th 6 weeks. For 10th grade, it was 6.3 during the 4th 6 weeks, 6.3 during the 5th 6 weeks, and 6.4 during the 6th 6 weeks. 10th grade experienced a minor increase during the 6th 6 weeks. For 11th grade, it was 6.2 during the 4th 6 weeks, 6.3 during the 5th 6 weeks, and 6.5 during the 6th 6 weeks. 11th grade experienced minor increases throughout the semester. For 12th grade, it was 6.2 during the 4th 6 weeks, 6.2 during the 5th 6 weeks, and 6.3 during the 6th 6 weeks. 12th grade experienced a consistent wake-up time, with a minor increase during the 6th 6 weeks. Numerically, 9th grade had the earliest wake-up time. Numerically, 11th grade had the latest wake-up time. Spending the most time on extracurricular activities and homework, the lowest average mood, the highest stress level, and getting the least amount of sleep is associated with having a later wake-up time, which applies for 11th grade. Spending the least amount of time on homework, having the most free time, the highest average mood, and getting the most amount of sleep is associated with an earlier wake-up time, which applies for 9th grade.
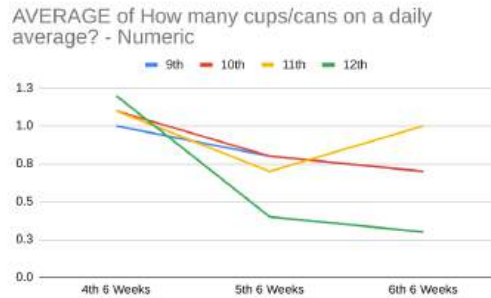
Fig 12: Numeric of average of how many cups/cans of caffeine consumed on a daily average throughout the semester by grades 9-12

Fig 13: Numeric of average of how many cups/cans of caffeine consumed on a daily average during the 4th 6 weeks by grades 9-12
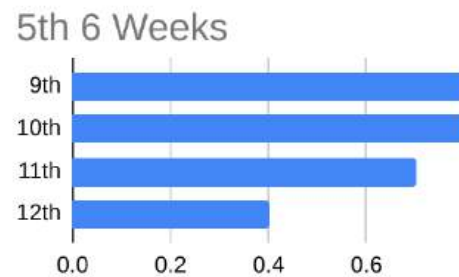
Fig 14: Numeric of average of how many cups/cans of caffeine consumed on a daily average during the 5th 6 weeks by grades 9-12
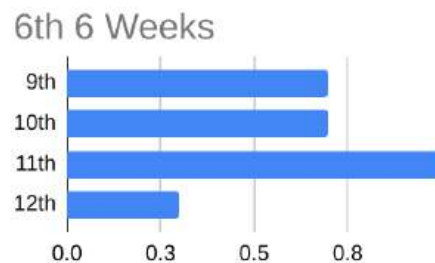
Fig 15: Numeric of average of how many cups/cans of caffeine consumed on a daily average during the 6th 6 weeks by grades 9-12

From the 4th 6 weeks to the 6th 6 weeks, the average of how many cups/cans of caffeine consumed on a daily average decreased across most grades, with a fluctuation for 11th grade. For 9th grade, it was 1.0 during the 4th 6 weeks, 0.8 during the 5th 6 weeks, and 0.7 during the 6th 6 weeks. 9th grade experienced a decrease throughout the semester. For 10th grade, it was 1.1 during the 4th 6 weeks, 0.8 during the 5th 6 weeks, and 0.7 during the 6th 6 weeks. 10th grade experienced a decrease throughout the semester. For 11th grade, it was 1.1 during the 4th 6 weeks, 0.7 during the 5th 6 weeks, and 1.0 during the 6th 6 weeks. 11th grade experienced a drop then an increase. For 12th grade, it was 1.2 during the 4th 6 weeks, 0.4 during the 5th 6 weeks, and 0.3 during the 6th 6 weeks. 12th grade experienced a decrease throughout the semester. Numerically, 12th grade drank the most amount of caffeine during the beginning of the semester and the least at the end of the semester.
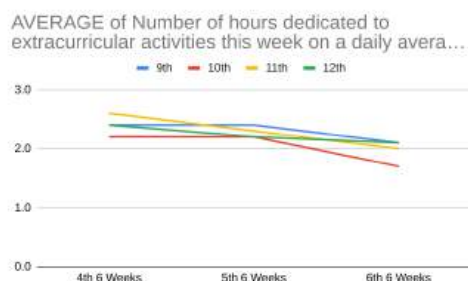
Fig 16: Numeric of average of number of hours dedicated to extracurricular activities per week on a daily average throughout the semester across grades 9-12

Fig 17: Numeric of average of number of hours dedicated to extracurricular activities per week on a daily average during the 4th 6 weeks across grades 9-12

Fig 18: Numeric of average of average of number of hours dedicated to extracurricular activities per week on a daily average during the 5th 6 weeks across grades 9-12
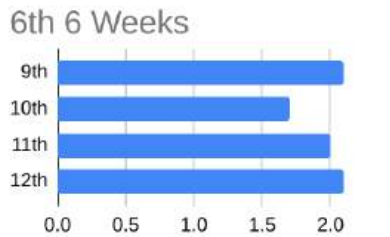
Fig 19: Numeric of average of average of number of hours dedicated to extracurricular activities per week on a daily average during the 6th 6 weeks across grades 9-12

From the 4th 6 weeks to the 6th 6 weeks, the average number of hours dedicated to extracurricular activities on a daily basis had either stayed consistent or dropped across all grade levels. For 9th grade, it was 2.4 during the 4th 6 weeks, 2.4 during the 5th 6 weeks, and 2.1 during the 6th 6 weeks. 9th grade experienced a consistent average then a drop. For 10th grade, it was 2.2 during the 4th 6 weeks, 2.2 during the 5th 6 weeks, and 1.7 during the 6th 6 weeks. 10th grade experienced a consistent average then a drop. For 11th grade, it was 2.6 during the 4th 6 weeks, 2.3 during the 5th 6 weeks, and 2.0 during the 6th 6 weeks. 11th grade experienced decreases throughout the semester. For 12th grade, it was 2.4 during the 4th 6 weeks, 2.2 during the 5th 6 weeks, and 2.1 during the 6th 6 weeks. 12th grade experienced a drop throughout the semester. Numerically, 11th grade spent the most time on extracurricular activities. Numerically, 10th grade spent the least amount of time on extracurricular activities.
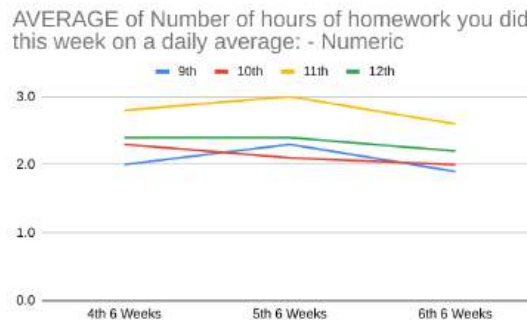
Fig 20: Numeric of average of number of hours of homework done per week on a daily average throughout the semester across grades 9-12
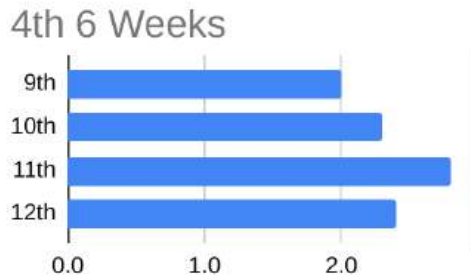
Fig 21: Numeric of average of number of hours of homework done per week on a daily average during the 4th 6 weeks across grades 9-12
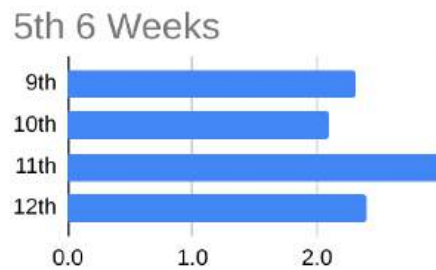
Fig 22: Numeric of average of number of hours of homework done per week on a daily average during the 5th 6 weeks across grades 9-12
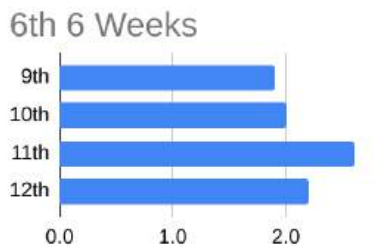
Fig 23: Numeric of average of number of hours of homework done per week on a daily average during the 6th 6 weeks across grades 9-12

From the 4th 6 weeks to the 6th 6 weeks, the average number of hours of homework done per week on a daily average has fluctuated based on grade. For 9th grade, it was 2.0 during the 4th 6 weeks, 2.3 during the 5th 6 weeks, and 1.9 during the 6th 6 weeks. 9th grade experienced an increase then a decrease. For 10th grade, it was 2.3 during the 4th 6 weeks, 2.1 during the 5th 6 weeks, and 2.0 during the 6th 6 weeks. 10th grade experienced a decrease throughout the semester. For 11th grade, it was 2.8 during the 4th 6 weeks, 3.0 during the 5th 6 weeks, and 2.6 during the 6th 6 weeks. 11th grade experienced an increase then a decrease. For 12th grade, it was 2.4 during the 4th 6 weeks, 2.4 during the 5th 6 weeks, and 2.2 during the 6th 6 weeks. Numerically, 11th grade spent the most amount of time on homework. In fact, 11th grade spent

the most amount of time on homework compared to any other grade throughout the semester. Numerically, 9th grade spent the least amount of time on homework.
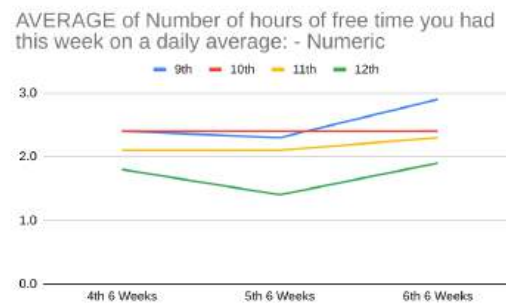
Fig 24: Numeric of average of number of hours of free time per week on a daily average throughout the semester across grades 9-12
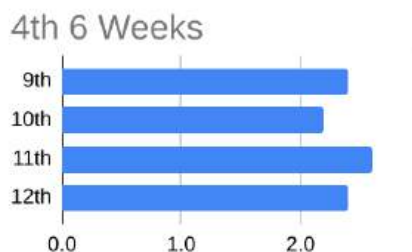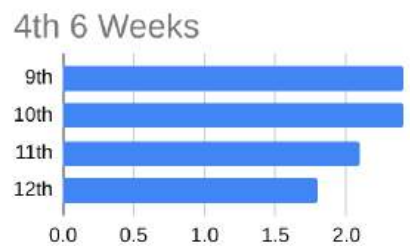

Fig 25: Numeric of average of number of hours of free time per week on a daily average during the 4th 6 weeks across grades 9-12
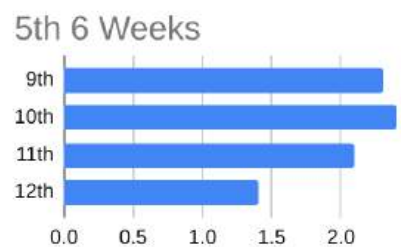
Fig 26: Numeric of average of number of hours of free time per week on a daily average during the 5th 6 weeks across grades 9-12
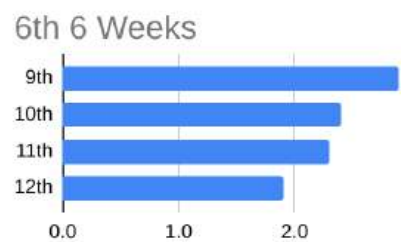
Fig 27: Numeric of average of number of hours of free time per week on a daily average during the 6th 6 weeks across grades 9-12

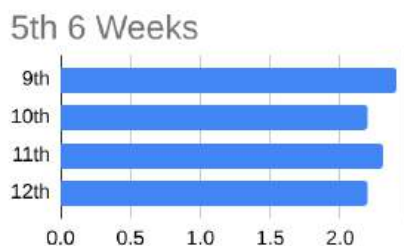From the 4th 6 weeks to the 6th 6 weeks, the average number of hours of free time per week on a daily average has stayed consistent, increased, or decreased based on grade. For 9th grade, it was 2.4 during the 4th 6 weeks, 2.3 during the 5th 6 weeks, and 2.9 during the 6th 6 weeks. 9th grade experienced a decrease then an increase. For 10th grade, it was 2.4 during the 4th 6 weeks, 2.4 during the 5th 6 weeks, and 2.4 during the 6th 6 weeks. 10th grade experienced a consistent amount of free time. For 11th grade, it was 2.1 during the 4th 6 weeks, 2.1 during the 5th 6 weeks, and 2.3 during the 6th 6 weeks. 11th grade experienced a consistent amount of free time followed with a slight increase. For 12th grade, it was 1.8 during the 4th 6 weeks, 1.4 during the 5th 6 weeks, and 1.9 during the 6th 6 weeks. 12th grade experienced a decrease then an increase. Numerically, 9th grade had the most amount of free time. Numerically, 12th grade had the least amount of free time.

Fig 28: Numeric of average productivity levels throughout the semester across grades 9-12
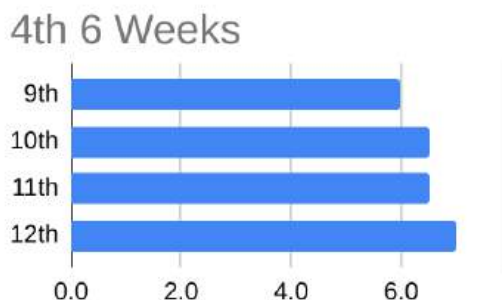
Fig 29: Numeric of average productivity levels during the 4th 6 weeks across grades 9-12
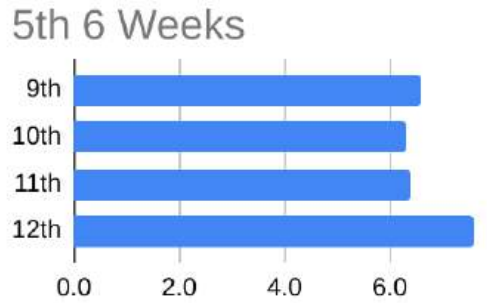
Fig 30: Numeric of average productivity levels during the 5th 6 weeks across grades 9-12
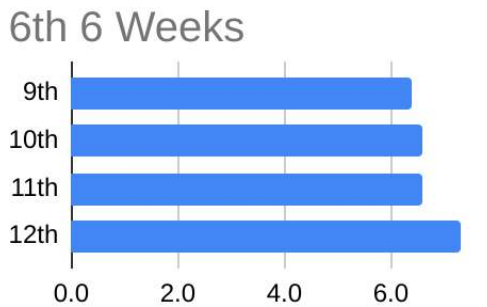
Fig 31: Numeric of average productivity levels during the 6th 6 weeks across grades 9-12
From the 4th 6 weeks to the 6th 6 weeks, the average productivity level fluctuated based on grade. For 9th grade, it was 6.0 during the 4th 6 weeks, 6.6 during the 5th 6 weeks, and 6.4 during the 6th 6 weeks. 9th grade experienced a significant increase then a slight decrease. For 10th grade, it was 6.5 during the 4th 6 weeks, 6.3 during the 5th 6 weeks and 6.6 during the 6th 6 weeks. 10th grade experienced a slight decrease then increase. For 11th grade, it was 6.5 during the 4th 6 weeks, 6.4 during the 5th 6 weeks, and 6.6 during the 6th 6 weeks. 11th grade experienced a slight decrease then increase. For 12th grade, it was 7.0 during the 4th 6 weeks, 7.6 during the 5th 6 weeks, and 7.3 during the 6th 6 weeks. 12th grade experienced a significant increase then a slight decrease. Numerically, 12th grade had the highest average productivity levels. Numerically, 10th grade had the lowest productivity levels.
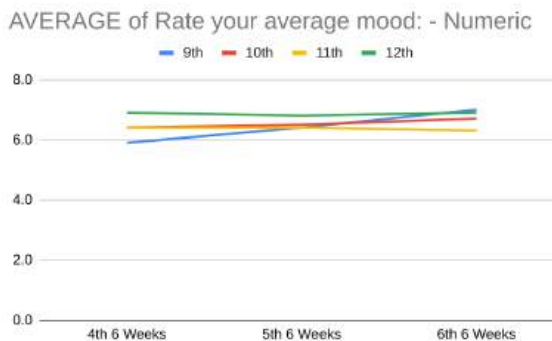
Fig 32: Numeric of average mood levels throughout the semester across grades 9-12
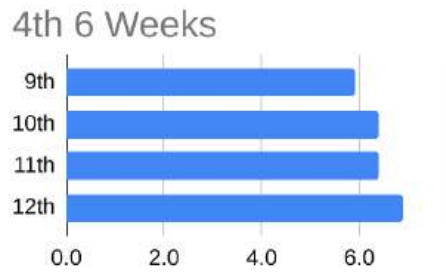
4th 6 Weeks

Fig 33: Numeric of average mood levels during the 4th 6 weeks across grades 9-12
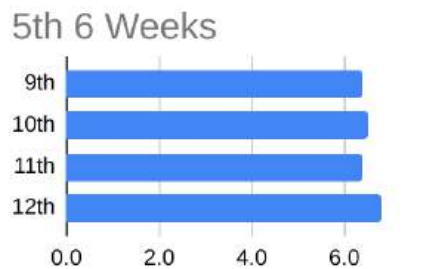
5th 6 Weeks

Fig 34: Numeric of average mood levels during the 5th 6 weeks across grades 9-12
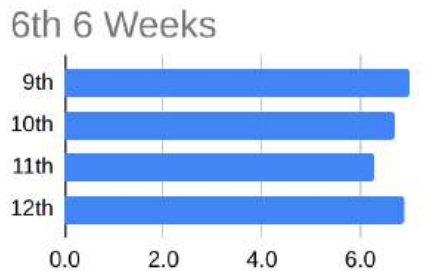
6th 6 Weeks

Fig 35: Numeric of average mood levels during the 6th 6 weeks across grades 9-12
From the 4th 6 weeks to the 6th 6 weeks, the average mood levels fluctuated based on grade. For 9th grade, it was 5.9 during the 4th 6 weeks, 6.4 during the 5th 6 weeks, and 7.0 during the 6th 6 weeks. 9th grade experienced increases throughout the semester. For 10th grade, it was 6.4 during the 4th 6 weeks, 6.5 during the 5th 6 weeks, and 6.7 during the 6th 6 weeks. 10th grade experienced slight increases throughout the semester. For 11th grade, it was 6.4 during the 4th 6 weeks, 6.4 during the 5th 6 weeks, and 6.3 during the 6th 6 weeks. 11th grade experienced a consistent average then a slight drop. For 12th grade, it was 6.9 during the 4th 6 weeks, 6.8 during the 5th 6 weeks, and 6.9 during the 6th 6 weeks. 12th grade experienced a slight drop then a slight increase. Numerically, 9th grade had the highest average mood. Numerically, 11th grade had the lowest average mood.
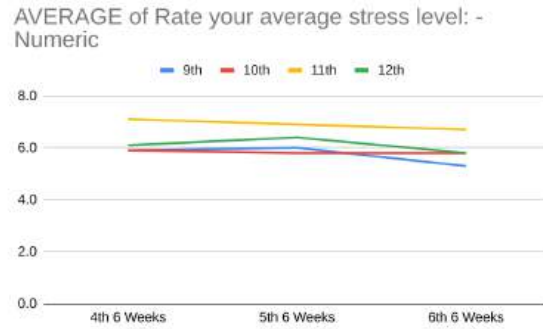
Fig 36: Numeric of average stress levels throughout the semester across grades 9-12
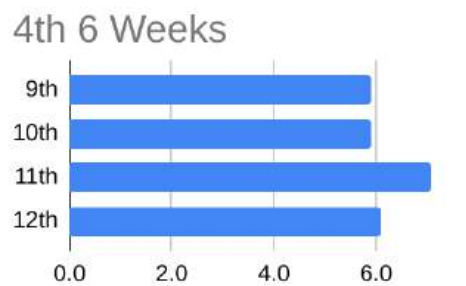
Fig 37: Numeric of average stress levels throughout the 4th 6 weeks across grades 9-12
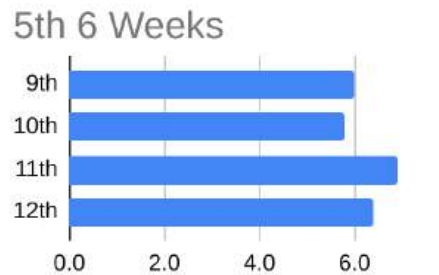
Fig 38: Numeric of average stress levels during the 5th 6 weeks across grades 9-12
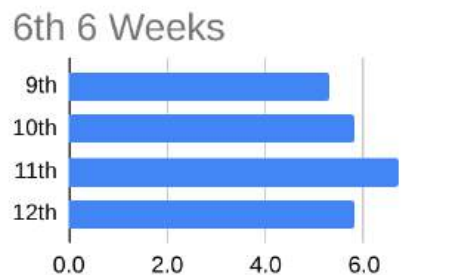
Fig 39: Numeric of average stress levels during the 6th 6 weeks across grades 9-12

From the 4th 6 weeks to the 6th 6 weeks, the average stress level fluctuated based on grade. For 9th grade, it was 5.9 during the 4th 6 weeks, 6.0 during the 5th 6 weeks, and 5.3 during the 6th 6

weeks. 9th grade experienced a slight increase then a significant drop. For 10th grade, it was 5.9 during the 4th 6 weeks, 5.8 during the 5th 6 weeks, and 5.8 during the 6th 6 weeks. 10th grade experienced a slight drop then a consistent average stress level. For 11th grade, it was 7.1 during the 4th 6 weeks, 6.9 during the 5th 6 weeks, and 6.7 during the 6th 6 weeks. 11th grade experienced drops of 0.2 per 6 weeks. For 12th grade, it was 6.1 during the 4th 6 weeks, 6.4 during the 5th 6 weeks, and 5.8 during the 6th 6 weeks. 12th grade experienced an increase then a decrease. Numerically, 11th grade reported the highest stress levels. Numerically, 9th grade reported the lowest stress levels.
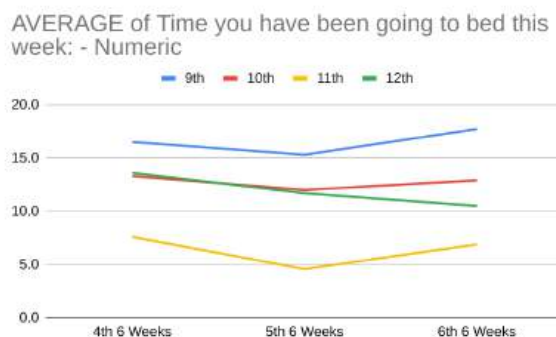
Fig 40: Numeric of average time students have been going to bed per week throughout the semester across grades 9-12
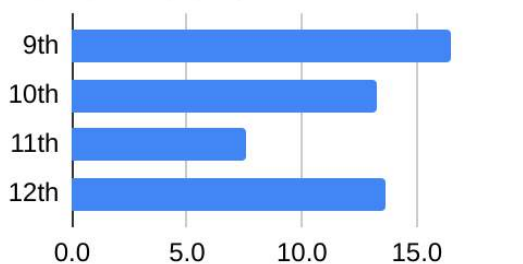
Fig 41: Numeric of average time students have been going to bed per week during the 4th 6 weeks across grades 9-12
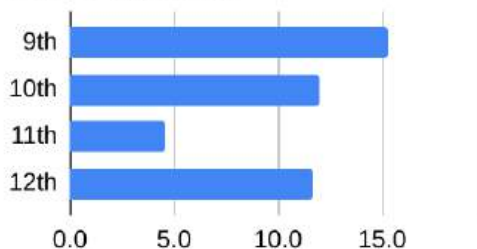
Fig 42: Numeric of average time students have been going to bed per week during the 5th 6 weeks across grades 9-12
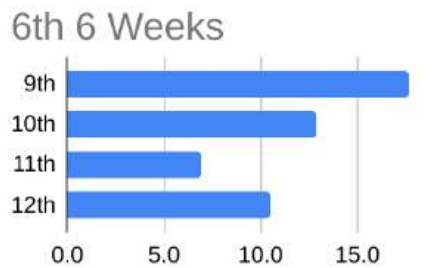
Fig 43: Numeric of average time students have been going to bed per week during the 6th 6 weeks across grades 9-12

From the 4th 6 weeks to the 6th 6 weeks, average bedtime per week has fluctuated based on grade. For 9th grade, it was 16.5 hours during the 4th 6 weeks, 15.3 hours during the 5th 6 weeks, and 17.7 hours during the 6th 6 weeks. 9th grade experienced a decrease then an increase. For 10th grade, it was 13.3 hours during the 4th 6 weeks, 12.0 hours during the 5th 6 weeks, and 12.9 hours during the 6th 6 weeks. 10th grade experienced a decrease then an increase. For 11th grade, it was 7.6 hours during the 4th 6 weeks, 4.6 hours during the 5th 6 weeks, and 6.9 hours during the 6th 6 weeks. 11th grade experienced a decrease then an increase. For 12th grade, it was 13.6 hours during the 4th 6 weeks, 11.7 hours during the 5th 6 weeks, and 10.5 hours during the 6th 6 weeks. 12th grade experienced a decrease throughout the semester. Grades may have extremely early or late sleep times because numerous students sleep after 12:00am. Sleep times were calculated based on military time.

Fig 44: Numeric of average of number of hours of sleep students have been getting per week throughout the semester across grades 9-12
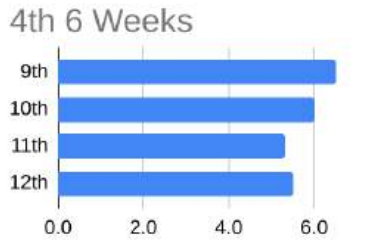
Fig 45: Numeric of average of number of hours of sleep students have been getting per week during the 4th 6 weeks across grades 9-12
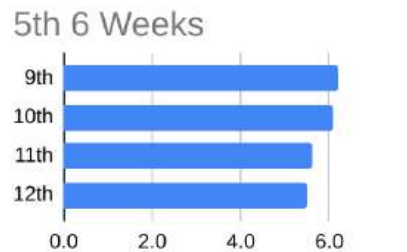
Fig 46: Numeric of average of number of hours of sleep students have been getting per week during the 5th 6 weeks across grades 9-12
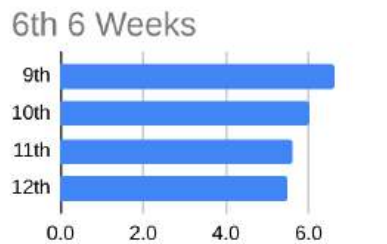
Fig 47: Numeric of average of number of hours of sleep students have been getting per week during the 6th 6 weeks across grades 9-12

From the 4th 6 weeks to the 6th 6 weeks, the average number of hours of sleep has fluctuated or stayed consistent based on grade. For 9th grade, it was 6.5 during the 4th 6 weeks, 6.2 during the 5th 6 weeks, and 6.6 during the 6th 6 weeks. 9th grade experienced a decrease then an increase. For 10th grade, it was 6.0 during the 4th 6 weeks, 6.1 during the 5th 6 weeks, and 6.0 during the 6th 6 weeks. 10th grade experienced a slight increase then a decrease to the original, 6.0. For 11th grade, it was 5.3 during the 4th 6 weeks, 5.6 during the 5th 6 weeks, and 5.6 during the 6th 6 weeks. 11th grade experienced a slight increase then remained consistent. For 12th grade, it was 5.5 during the 4th 6 weeks, 5.5 during the 5th 6 weeks, and 5.5 during the 6th 6 weeks. 12th grade experienced a consistent amount of sleep throughout the semester. Numerically, 9th grade got the most amount of sleep. Numerically, 11th grade got the least amount of sleep.

Qualitative Results

Majority of students reported schoolwork as their top factor affecting stress levels and productivity levels. After schoolwork came extracurricular activities, followed by work and personal issues. During the 4th 6 weeks, the majority of students, 35.0%, reported that they did not feel well-rested when waking up throughout the week. 51.9% reported drinking caffeine. 63.6% reported getting exercise, which made 56.4% feel better. 94.4% did have social interactions, which made 82.8% feel better. 70.3% completed all schoolwork. 35.8% reported feeling focused. 96.7% reported not having a sleep disorder, and 3.3% of students reported having insomnia. During the 5th 6 weeks, the majority of students, 36.7% and 36.7%, reported feeling well-rested and feeling somewhat well-rested when waking up throughout the week. 51.9% reported not drinking caffeine. 64.0% reported getting exercise, which made 57.4% feel better. 96.9% reported having social interactions, which made 87.4% feel better. 73.1% reported finishing all schoolwork. 42.9% reported feeling like they could focus. 95.0% reported having no sleep disorder, 3.3% reported having insomnia, and 1.7% reported having POTS. During the 6th 6 weeks, the majority of students, 42.9%, reported feeling well-rested when waking up. 50.5% reported not drinking caffeine. 57.9% reported getting exercise, which made 56.2% feel better. 93.1% reported having social interactions, which made 82.6% feel better. 79.8% reported finishing all schoolwork. 48.8% reported feeling like they could focus. 95.0% reported not having a sleep disorder, 3.3% reported having insomnia, and 1.7% reported having POTS.

Discussion

The patterns show that students who sleep more hours per night have less stress factors, a lower stress level, higher average mood, and report an overall better well-being. The results came just as expected. Throughout the semester, each week, each student reported their average stress levels, average productivity levels, average time spent on homework, etc. Trends have been observed between the variables. The earliest wake-up time is associated with spending the least amount of time on homework, having the most free time, having the highest average mood, having the lowest stress levels, and getting the most sleep, which applies specifically and only to 9th grade. Spending the least amount of time on extracurricular activities is associated with having the lowest productivity levels, which applied specifically and only to 10th grade. The latest wake-up time is associated with spending the most amount of time on extracurricular activities, spending the most amount of time on homework, having the lowest average mood, having the highest stress levels, and getting the least amount of sleep, which applies specifically and only to 11th grade. Drinking the most amount of caffeine at the beginning of the semester then the least at the end of the semester is associated with having the least amount of free time and having the highest productivity levels, which applies specifically to 12th grade. There are significantly large differences between grades. For high school students, sleep is extremely low, and stress is extremely high. On average by grade, 0% of grades reported getting the recommended 8-10 hours of sleep essential for development per night. This data supports the idea that school should start later in the morning.

Limitations

        There are limitations to this research project. There were 60 total participants, who reported their own data. Late submissions were accepted, and self-reporting data, among other factors, may create a small margin of error. Qualitative data was limited to a few categories per question. Even though students from various parts of the United States of America participated and only complete data with no gaps was accepted, this data may not apply to every student and community in the United States of America. Every student is unique in their academic and extracurricular pursuits.

Conclusion

        The hypothesis was that younger students will get more sleep each night and therefore, have an overall better mood, performance, health, etc. Students who sleep more on average will have an overall better mood, performance, health, etc. The hypothesis was accepted. 9th grade students reported overall better sleep, mood, stress, and less time devoted to schoolwork. However, 11th grade students reported overall worse sleep, mood, stress, and more time devoted to schoolwork. More time devoted to schoolwork has a direct correlation with overall worse average well-being, well-being defined as better sleep times, lower stress levels, higher average mood, etc. Students who reported less time devoted to schoolwork outside of school reported a better overall performance, as proved by the data. Future research may be conducted on how students' well-being improves with a lighter schoolwork load. The data presented the level of stress and struggle students experience during the school year, which is extremely high for an age range of 14-18. This data may present schools with reasons to push back start times in the morning, reduce the amount of workload on students outside of school, and take their students' well-being more into consideration, which will ultimately lead to better overall performance. The data shows just how important well-being is to achieve a better performance. Students, schools, and parents should consider working together to make well-being for all high school students a reality while working to succeed at the same time.

Acknowledgements

        I am immensely grateful to my family for their support during this study. Their support has made my work possible. I would like to especially thank my mother, who supported me at times I doubted myself. She was the person who encouraged me to keep going. I would also like to thank all 60 of the participants who devoted their time and energy towards making this study a success. Without these 60 individuals, this research study would not have been possible. Thank you, from the bottom of my heart.

Works Cited

        No external resources were referenced. All data and findings come directly from polls filled out by participants.

Exploring the Relationship Between Consumers' Brand Attitude and Their Purchasing Intention Regarding Sustainability Claims By Daneya Chaudhry

Abstract

Due to growing concerns about global warming and climate change, consumers are becoming more conscious of making environmentally responsible choices. Many businesses attract eco-conscious consumers by using sustainability claims such as "eco-friendly," "fair trade," or "carbon neutral", which help them directly communicate their values to potential buyers. This study explored the relationship between sustainability cues, brand attitude, and purchase intention. The study dove deeper into how consumers perceive a brand with sustainability cues and whether a positive attitude subsequently translates into stronger purchase intention.

Data was collected online via the use of a Google Form. The brand, Alter Eco, which is known for its sustainable and fair-trade edibles, was used as the brand stimulus. 40 participants (7 males and 33 females), aged 16 to 68 years (mean age of 38 years), completed the survey. Measures used were the Brand Attitude Scale (Lutz 49-59) and the Purchase Intention Scale (Spears & Singh 53-66). Results showed above-average scores for both brand attitude (M = 22.73, SD = 5.14) and purchase intention (M = 22.50, SD = 6.00). A strong positive correlation was found between brand attitude and purchase intention (r = 0.824), which shows that positive brand attitude is related to stronger buying intentions. Consumers' interest in the product was also positively related to their confidence in it. (r = 0.69). These findings suggest that sustainability cues can influence brand attitudes and, consequently, their purchase intentions. This highlights the growing importance of sustainability branding and how it impacts consumer behaviour in today's markets.

Keywords

Brand Attitude, Purchase Intention, Sustainability Claims, Consumer Perception, Sustainable Packaging

1. Introduction

Studying sustainability claims have become increasingly important in today's world as consumers are becoming increasingly cautious about global issues such as climate change, plastic pollution, and unethical labor practices.

Sustainability claims are statements made by companies to indicate that their products or practices are environmentally friendly, socially responsible, or ethically produced. These sustainability claims include terms such as "plastic free," "carbon neutral," or "recyclable". According to Frey et al, grocery and pharmacy producers are stocked up with products which are claimed to be "environmentally sustainable," "eco-friendly," "fair trade," and have similar designations, which stresses how common such claims have become in Consumer Packaged Goods (CPG) settings. Such cues can help promote a business by helping them communicate

their principles and also make consumers feel that their purchasing decisions contribute to a greater social and environmental impact.

The way consumers respond to these claims is known as perception. Perception is shaped by personal experiences, cultural values, beliefs, and biases, influencing how individuals interpret and react to information. It plays a major role in purchase decisions. For instance, while some consumers may view sustainability claims as genuine, meaningful, and important, others may see them as exaggerated marketing tactics.

Another key concept is brand attitude, which refers to positive or negative feelings consumers form based on their interactions with a particular brand. When a brand presents itself as "sustainable", more consumers are likely to develop a positive attitude or feeling about the brand and are more likely to prefer its products over competitors.

Brand attitude has been defined by Mitchell and Olson as an *"individual's internal evaluation of the brand."*(Mitchell and Olson 318) Giner-Sorolla further expands on this definition explaining two distinctive features of attitude, which have mostly been constant over the years; firstly, it is directed towards an object, (in this case, a brand)(Giner-Sorolla 441-461) Secondly, attitude is based on evaluating something, i.e., there is *"imputation of some degree of goodness or badness"* to the attitudinal object. (Eagly and Chaiken 459-466) Lastly, the meaning of internal evaluation, derived by Mitchell and Olson's, is too significant. This conveys that it is an internal condition. On the other hand, attitude is a state *"that endures for at least a short period of time and presumably energizes and directs behavior."* Eagly and Chaiken add that attitude is a permanent condition. (Eagly and Chaiken 459-466) With reference to Spears & Singh's attitude toward the brand is a relatively lasting, one-sided, concise evaluation of the brand that possibly promotes behavior. (Spears & Singh 53-66)

Theory of Reasoned Action (TRA) (Fishbein and Ajzen) and Theory of Planned Behavior (TPB) (Ajzen 179-211) are two closely related theories that explain how attitudes influence behaviour. They suggest that a person's intention to perform a behaviour (behaviour intention), is determined by their attitude towards the actions and the social expectations associated with it.. Subjective norms arise from a person's social and environmental surroundings and reflect the perceived influence on their behavior. Generally, perceived control and the impact of intention on changes in behaviour can be enhanced by positive attitudes and subjective norms. Inferring, if an individual has a favourable attitude towards a specific object, their intention to engage with or buy that object is likely to increase. Hence, if an individual has a positive brand attitude, it is more likely that they will buy or consume that product.

Purchase intention refers to a stage between a consumer and a seller when the consumer is willing to enter into a purchase agreement with the seller. It is defined as the degree to which a consumer is inclined to buy a product or service because they find it beneficial, or hold a positive attitude and opinion about it. (Tilahun, Berhan, and Tesfaye 50)

In the current study, the main aim is to explore how sustainability cues influence a person's attitude towards the brand, whether these cues lead to greater likeability for that brand or not. Further, the study aims to understand how this brand attitude contributes to an

individual's intention to actually purchase the product and whether a possible relationship exists between the two. The research also examines whether people's preference for sustainability makes them more likely to consume products that display sustainability-related cues or logos.

2. Methodology

2.1 Research Design
This study used a quantitative research design to collect and analyse data. The study used chocolate truffles from Alter Eco as its brand stimulus for the participants. Alter Eco is known for its sustainable and fair trade products, focusing on creating food that is safe for the planet and human health. Their packaging and promotional materials clearly display sustainability cues, making their brand suitable for this research.

2.2 Tools Used
For this study, a standardised questionnaire was used. The questionnaire was created using Google Forms and included 15 questions divided into 4 sections.
The first section covered demographic details such as name (initials only), age, gender, employment status, and factors that influence how much an individual trusts a brand's sustainability claims. The next three sections of the questionnaire are described below:

2.2.1 Brand Stimulus
Alter Eco is a brand that focuses mainly on chocolate-based food products. The brand supports Fairtrade, uses climate-neutral packaging, emphasizes recycling, and uses compostable wrappers. For this research, images of Alter Eco's chocolate truffles displaying these claims were shown to the participants. This brand was chosen as Alter Eco is not a well-known brand in India, which helped reduce bias or pre-existing perceptions among participants.

2.2.2 Brand Attitude Scale
The Brand Attitude Scale (BAS), developed by Richard Lutz in 1975, was used to measure participants' attitudes toward the brand. It is a self report tool which uses the cognitive components of brand attitude through 6 items rated on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree). To assess the BAS, firstly, each item's score is summed. The Brand Attitude Scale's score has a possible range from 6- 42. Higher scores reflect a more positive brand attitude. This scale is well established, valid, and widely used in marketing research

2.2.3 Purchase Intention Scale
The Purchase Intention Scale, developed by Spears and Singh in 2004, was used to measure consumers' intention to buy the product with a Cronbach's alpha of 0.97. It is a tool with 5 items with a 7-point semantic differential scale. Examples of bipolar adjective pairings

include "Never / Definitely, Definitely do not intend to buy / Definitely intend to buy, Very low / Very high purchase interest, Definitely not buy it / Definitely buy it, Probably not buy it / Probably buy it". Displaying contrasting features, it helps the buyer determine the likelihood of buying a product or service. The consumer selects a point on the scale to represent how much they agree with the statement.

2.3 Sample

A convenience sampling method was employed to collect data from the participants. The total sample consisted of 40 participants, wherein it included 7 Males and 33 females. The participants' ages varied from 16- 60+, with a mean age being 38. In the sample, 25 participants were employed and 15 were unemployed. The sample represented a mix of age groups and employment statuses, providing diverse perspectives on sustainability and brand attitude.

2.4 Informed Consent and Ethical Considerations

Before participating, individuals provided digital informed consent. They were informed that participation was voluntary and if at any point during the study they wished to withdraw their consent, they may do so. The participants' confidentiality was maintained, precaution was taken, such as taking their initials instead of full names, and data was kept private. The research adhered to the ethical considerations required for the data collection and research.

2.5 Data Collection & Analysis

Data collection was primarily quantitative, collected online using a Google Form. The population holding the purchasing power was asked to participate in the study. The survey included images of the brand stimulus to help assess the relation of sustainability cues with brand attitude and purchase intention. Quantitative analysis was carried out for both descriptive and inferential statistics. Furthermore, the data was analysed by calculating the mean and the standard deviation for age, brand attitude scale, and purchase intention. The data was analysed for the understanding of what factors contribute to people's perceptions. Lastly, a correlation analysis was conducted to examine the link between purchase intention and brand attitude.

3. Results and Discussion

The research study aimed to examine the relationship between brand attitude and purchase intention among consumers. The sample size included 40 participants in the age range of 16- 68 years, with a mean age of 38. The sample included 7 male and 33 female participants (Fig. 1), with 25 participants currently employed and 15 participants not employed (Fig. 2). The sample is diverse as it has responses from a wide range of age groups, a representation of working and non-working consumers, and responses from both males and females.
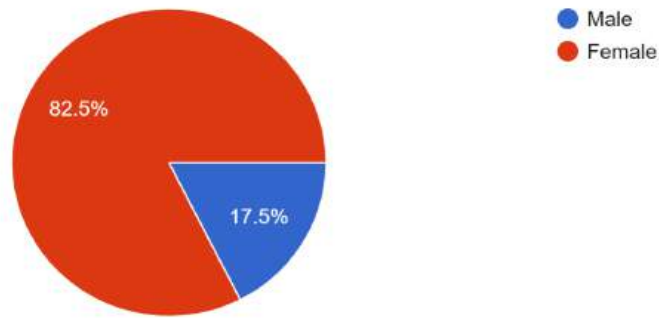
Figure 1: Gender Distribution of the participants in the study
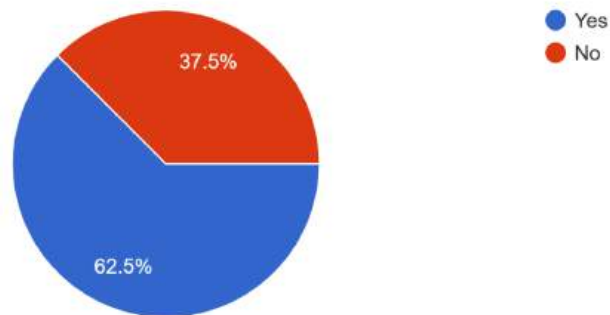
Figure 2: Employment status of the participants in the study

In Figure 2, the employment status of the participants is recorded. "Yes" means that the participant is employed, and "No" means that the participant is not employed.
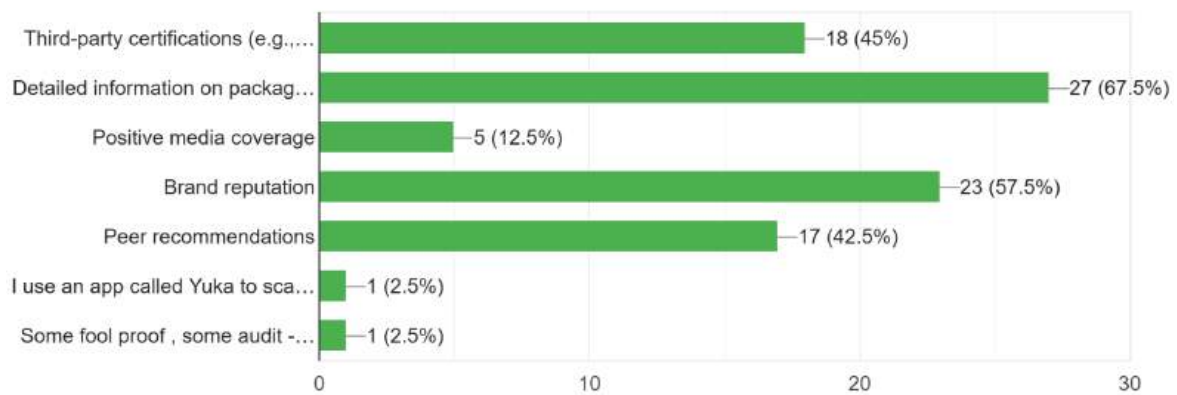
Figure 3: Factors that would make participants trust a brand's sustainability claim more

In Figure 3, it is observed that 'detailed information on packaging/website' has the highest number of responses with 27 (67.5%) participants voting for it, lowest is 'Positive media coverage' with 5 (12.5%) participant responses. Other factors like Brand reputation had 23 (57.5%) participant responses, then Third-party certifications (e.g., Fairtrade, USDA Organic) which had 18 (45%) participant responses, and lastly peer recommendation which had 17 (42.5%) participant responses.

Table 1: Mean and Standard deviation of Brand attitude and purchase intention

Variable	Mean	Standard Deviation
Brand Attitude	22.73	5.14
Purchase Intention	22.50	6.00

Table 1 shows that the brand attitude has M(SD)= 22.73 (5.14), indicating that participants demonstrated above-average brand attitude for the product displayed in the stimulus. Similarly, Purchase intention M(SD)= 22.50 (6), suggesting that participants also showed an above-average purchase intention for the product displayed in the stimulus.

Majer et al. conducted a systematic review on the effects of visual sustainability labels on consumer perception and behavior, which found that sustainable labels and packaging cues do significantly impact consumers' brand attitude and purchase intention. (Majer et al. 1-14) It is mostly the label that has the most impact on the attitude of the consumers as it provides utility, increases willingness to pay, and has a significant impact on the behaviour.

Table 2: Correlation between consumers' brand attitude and purchase intention

Variables	Brand attitude	Purchase intention
Brand attitude	1	0.824*
Purchase intention	0.824*	1

In Table 2, it can be seen that there is a strong positive correlation between brand attitude and purchase intention, with r(40)= 0.824. This signifies that the higher the brand attitude, the higher the purchase intention will be, and vice versa.

The results of the study by Gidaković et al. found that consumers have a better brand impression when it is perceived to be sustainable, they create a more positive attitude towards it, which increases the willingness of the consumers to purchase from that brand. (Gidaković et al. 556-568.) Similarly, in this study, where the brand stimulus has sustainability cues well displayed, there is a positive correlation between brand attitude and purchase intention. Both

studies emphasise that sustainability claims lead to better brand attitude and hence, stronger buying intentions.

Table 3: Correlation between consumers' confidence and interest in the product

Variables	Confidence in the product	Interest in the product
Confidence in the product	1	0.69*
Interest in the product	0.69*	1

Table 3 shows a positive correlation of r(40)= 0.69 between confidence in the product and interest in the product. This indicates that the higher the confidence, the higher the interest in the product and vice versa.

The results further reinforce the positive relationship between brand attitude and purchase intention of consumers indulging in products with sustainability claims. The research had many strengths, which included a diverse sample in terms of gender and age of the participants. It is one of the few exploratory studies examining brand attitude and purchase intention related to sustainability claims within the Urban Indian context. In the study, the ethical standards were met. However, using an already established brand as the stimulus may have led to participant bias due to pre-existing opinions about the brand. Future research can address this limitation by using an experimental design with a lesser-known brand to better understand the impact of sustainability claims.

4. Conclusion

The study aimed to explore the relationship between sustainability claims, brand attitude, and purchase intention among consumers. Based on the literature reviewed, it was predicted that stronger sustainability claims would positively influence a consumer's attitude toward the brand and, in turn, increase purchase intention. The results supported this, wherein there was above-average brand attitude and purchase intention reported by the participants for a brand stimulus with sustainability claims.

Further, a positive correlation between sustainability perception, brand attitude, and purchase intention was found, including a positive relationship between consumers' confidence and interest in the product. The study had a few strengths and limitations as discussed above, and future research can aim to bridge the experimental gap.

Works Cited

Ajzen, Icek. "The Theory of Planned Behavior." Organizational Behavior and Human Decision
Processes, vol. 50, no. 2, 1991, pp. 179-211.

Eagly, Alice H., and Shelly Chaiken. The Psychology of Attitudes. Harcourt Brace Jovanovich
College Publishers, 1993.

Fishbein, Martin, and Icek Ajzen. Belief, Attitude, Intention, and Behavior: An Introduction to
Theory and Research. 1977.

Frey, S. B. A. J., et al. "Consumers care about sustainability—and back it up with their wallets."
McKinsey & Company, 2023.

Giderković, Petar, et al. "The influence of brand sustainability on purchase intentions: the
mediating role of brand impressions and brand attitudes." Journal of Brand Management,
vol. 29, no. 6, 2022, pp. 556-568.

Giner-Sorolla, Roger. "Affect in attitude." In Dual Process Theories in Social Psychology, 1999,
pp. 441-461.

Lutz, Richard J. "Changing Brand Attitudes through Modification of Cognitive Structure."
Journal of Consumer Research, vol. 1, no. 4, Mar. 1975, pp. 49-59.

Majer, Johann M., et al. "The effects of visual sustainability labels on consumer perception and
behavior: A systematic review of the empirical literature." Sustainable Production and
Consumption, vol. 33, 2022, pp. 1-14.

Mitchell, Andrew A., and Jerry C. Olson. "Are Product Attribute Beliefs the Only Mediator of
Advertising Effects on Brand Attitude?" Journal of Marketing Research, vol. 18, no. 3,
Aug. 1981, pp. 318-332.

Spears, Nancy, and Surendra N. Singh. "Measuring Attitude Toward the Brand and Purchase
Intentions." Journal of Current Issues & Research in Advertising, vol. 26, no. 2, 2004, pp.
53-66.

Tilahun, Mulatu, Eshetie Berhan, and Gezahegn Tesfaye. "Determinants of consumers' purchase
intention on digital business model platform: evidence from Ethiopia using partial least
square structural equation model (PLS-SEM) technique." Journal of Innovation and
Entrepreneurship, vol. 12, no. 1, 2023: 50.

The Mathematical Origin of Financial Chart Patterns By Srivatsa Susarapu

Abstract

This research investigates the structural origins of technical analysis patterns in financial markets, challenging the prevailing "Behavioral Finance" paradigm which attributes chart formations, such as Head and Shoulders, Wedges, and Double Bottoms, to collective human psychology. I suggest that these patterns are not psychological signals, but rather emergent geometric artifacts of stochastic price action constrained by volatility bounds.

To test this hypothesis, I developed the Krafer Model, a "Taker-Only" market engine that eliminates human agency, limit-order market makers, and order-book memory. The model utilizes a pure Bernoulli trial (p=0.5) for trade direction, with price velocity regulated by a 1% Absolute Bound to simulate market friction. The model, developed based on (and named after) a similar experiment run by Krafer; an independent Youtube Channel dedicated to pushing the boundaries of A.I. and statistics; is meant to eliminate almost all psychology from the simulation entirely.

By aggregating 15,000 discrete ticks into OHLC (Open, High, Low, Close) candlestick data across multiple sequential sessions, I observe the spontaneous formation of classic technical patterns. The results demonstrate that these "signals" appear consistently in a purely random environment, suggesting that the geometry of the market is a byproduct of mathematical variance and scale-invariant price movement. These findings imply that chart patterns are emergent constants of bounded price movement. Rather than psychological signals, they are the Internal Physics of the market skeleton, suggesting that the geometry traders' identity exists independently of the human agency traditionally thought to create it.

1.Introduction

For decades, since the advent of market analysis, the prevailing theory behind market movement has been split into two incompatible camps. The Efficient Market Hypothesis (referred to as EMH) posits that because all available information regarding a stock is already reflected in the price, the next move can only be caused by new information. Essentially, it asserts that it is impossible for an investor to consistently outperform or predict the market through any means unless they take on additional risk. (EBSCO Research)

The competing theory, Technical Analysis (abbreviated as TA), functions on the foundation that markets move in recognizable, repeatable patterns. Practitioners of TA argue that price action is not a series of independent events, but a reflection of the collective psyche of market participants. In this view, a "Head and Shoulders" pattern is more than a geometric coincidence; it is a visual representation of a shift from investor euphoria to exhaustion and eventual panic. (EBSCO Research)

1.1 The Conflict

The tension between EMH and TA creates a fundamental divide in financial literature. If EMH is correct, then technical patterns are merely a form of apophenia which is the human tendency to perceive meaningful patterns within random data. In this framework, the "Head and Shoulders" is no more significant than seeing a cloud that looks like an animal. Conversely, if TA is correct, then markets are not efficient, as past price patterns would allow for the prediction of future moves, thereby violating the "Random Walk" principle. This conflict has historically been a binary choice: either the patterns are real and driven by human emotion, or the patterns are fake and driven by random noise.

1.2 Patterns without People

This paper introduces the Krafer Model to bridge this gap, suggesting that patterns are not "put there" by news or "put there" by emotions; they are the inevitable byproduct of a bounded stochastic process. By utilizing a Taker-Only market engine, one stripped of human agency, market makers, and order-book memory, I can test a "Structural Null Hypothesis." I contend that if a purely random system produces these patterns, then the geometry of the market is a byproduct of mathematical variance and scale-invariant price movement, rather than collective human behavior. Essentially, there is no relationship between human psychology and market patterns.

It is important to note that the viability of Technical Analysis as a market strategy will not be the primary focus of this investigation. We are not concerned here with the efficacy of these patterns in predicting future returns or generating alpha. Rather, our focus is squarely on the causal origin of the patterns themselves. By adopting a structural framework, we posit that the 'Support' or 'Resistance' identified on a chart is a mathematical threshold where the engine's path-dependency creates a density of price action, regardless of external sentiment.

1.3 Research Background

Intuitively, the theory behind TA makes sense. A stock chart offers a display of different numerical human decisions. Humans tend to make decisions in patterns, therefore, one should be able to utilize said patterns to predict them in the future. However, critics are often quick to point out that predicting mass human psychology is impossible or at the very least incredibly difficult, and the patterns can only be consistently seen in hindsight.

Beyond the academic debate, the credibility of Technical Analysis is further complicated by its commercialization. Because many TA patterns are visually intuitive and easy to identify, they have frequently been co-opted by bad-faith actors. These individuals often peddle these patterns as "sure-shot" or "get-rich-quick" schemes, promising high-probability returns with minimal risk.

This saturation of pseudo-educational content has led many serious financial theorists to dismiss Technical Analysis entirely, viewing it as a tool for marketing rather than a tool for science. However, this dismissal may be premature. If we strip away the skepticism and variability, one question still remains: Why do they appear at all?

If the patterns are as simple as critics say, and as common as proponents say, then they should be reproducible in a controlled environment. By removing the Human Decision variable entirely, we can test whether these patterns are a product of predetermined success sold by gurus, or a fundamental property of market physics.

2.Literature Review

2.1 Mandelbrot and the Fractal Nature of Markets

In his seminal 1963 paper, *"The Variation of Certain Speculative Prices,"* and his later work, *The (Mis)behavior of Markets (2004)*, Mandelbrot launched a devastating critique of the "Mild Randomness" found in standard financial models. He argued that the traditional Bell Curve, the Gaussian distribution, is fundamentally ill-equipped to describe the reality of market price action.

The most profound implication of Mandelbrot's work is the concept of Scale Invariance. He observed that a chart of a commodity's price over 60 minutes often looks structurally identical to a chart of that same commodity over 60 weeks. This "Self-Similarity" suggests that the patterns identified by Technical Analysts, such as wedges, channels, and triangles, are not products of a specific human timeframe or "trading session psychology." Instead, they are the result of a Scaling Power Law. In Mandelbrot's view, the "jaggedness" or "roughness" of a price chart is a constant, regardless of the zoom level.

Mandelbrot introduced the "Joseph Effect" (referencing the biblical seven years of plenty and seven years of famine) to describe how price action clusters into trends. Unlike a standard coin flip where the result resets completely every time, market price action has "persistence."

While Mandelbrot's work focuses on the resulting "roughness" of the chart, the Krafer Model addresses the mechanical cause of that roughness through the 1% Absolute Bound. In a theoretical unconstrained random walk, a price could transition from $5.00 to $50.00 in a single discrete step. However, such a "teleportation" of value is not representative of real-world equity markets.

In actual exchange environments, price discovery is a process of navigating a Limit Order Book. For a stock to move from $5.00 to $50.00, it must first consume all available liquidity at $5.01, 5.02, and every cent in between. The 1% bound in our simulation acts as a proxy for this Market Friction.

By preventing large spikes and forcing the price to move incrementally (e.g., to $4.95 or $5.05 instead of $50.00), the model creates Path Dependency. This is the secret ingredient for chart patterns: if the price is forced to walk rather than teleport, it must necessarily form geometric structures like curves, trends, and consolidations. Without this bound, the chart would appear as white noise; a cloud of points with no visual connection. With the bound, the "Random Walk" becomes a "Physical Walk," giving rise to the very patterns Technical Analysts attempt to trade.

2.2 Behavioral Finance and Sentiment

With the mathematical baseline established by Mandelbrot and the mechanical baseline established by our price-continuity bounds, we must address the "Human Element." This is where the work of Smith, Wang, Wang, and Zychowicz (2014) becomes vital.

Smith et al. provide a critical reality check for any theory of technical analysis. They studied the hedge fund industry, perhaps the most "rational" and "informed" segment of the market, and found that nearly 20% of funds explicitly utilize TA. This suggests that price patterns are not merely "get-rich-quick" fantasies for retail traders, but are viewed as legitimate data points by institutional professionals. It is important to note that a main reason these professionals find success through TA is due to their back-testing strategies, as comparatively, an independent trader will be incapable of having the same verifiability.

The core finding of Smith et al. is that TA is most effective during periods of high market sentiment. They argue that when investors are emotional, they create "trends" and "bubbles" that TA practitioners can identify and exploit.

However, the Krafer Model challenges the *causality* of this finding. If our simulation, which has zero sentiment and zero human participants, can produce the same primary patterns that hedge funds trade, then sentiment may not be the *creator* of the pattern. Instead, sentiment may simply be a volatility multiplier. In this framework, humans don't "create" the patterns; the math of the bounded walk creates the pattern, and human emotion makes that pattern larger and more obvious.

2.3 Data Mining and the Illusion of Pattern

While Mandelbrot argues for the geometric reality of markets and Smith et al. provide evidence of institutional utility, a significant body of literature suggests that the "success" of technical analysis is an illusion. The most prominent modern critic is David Aronson, author of *Evidence-Based Technical Analysis (2006)*.

Aronson argues that if a researcher (or a trader) looks at enough random data, they will eventually find a pattern that *appears* to have predictive power purely by chance. He calls this Data Mining Bias. In the context of the Krafer Model, an Aronson-style critique would suggest that finding a pattern in 15,000 random ticks is not a discovery of market physics, but rather a mathematical inevitability of looking at a large enough dataset.

Aronson posits that the human brain is so desperate for order that it ignores the thousands of "failed" patterns and only remembers the "successful" ones. This Selection Bias is what gives TA its unearned reputation. Essentially, patterns are only visible because we want them to be.

2.4 Lo, Mamaysky, and Wang: The Quest for Verifiability

Before the Krafer Model can examine the *cause* of patterns, it must acknowledge the landmark study that first proved their *existence* through automation. In their 2000 paper, *"Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and*

Securities Pricing," Lo, Mamaysky, and Wang (often called the "Big Brother" paper of automated TA) moved technical analysis from the realm of "art" to the realm of "science."

Using Kernel Regression, the authors developed a systematic way to identify patterns like Head-and-Shoulders and Double-Tops without human "eyeballing." Their findings were revolutionary: they discovered that these patterns occur far more frequently in real-market data than in a standard, unconstrained random walk. Crucially, they found that these patterns do offer a degree of incremental predictive power, suggesting that TA is not entirely a "pseudoscience."

However, Lo et al. attributed this "predictive edge" to market psychology and human learning. They posited that patterns are the visual signatures of humans reacting to information. The Krafer Model utilizes the automated detection methods inspired by Lo et al. but applies them to a Sentiment-Zero environment. If the same geometric success observed by "Big Brother" occurs in our simulation, it suggests that Lo et al. may have correctly identified the *existence* of patterns but potentially misattributed their *origin* to human behavior rather than bounded price movement.

Additionally, though a truly valuable paper with a broad depth of research, the experiment itself is quite dated, and computer algorithms have come a long way in the past quarter-century. Lo et al.'s conclusion that "Technical Analysis may well be the next frontier for such methods" holds incredible weight, especially within the context of today's computing power.

2.5 Modern Media and the "Retail Perspective"

While academic journals provide the mathematical foundation, the cultural spread of Technical Analysis often occurs through modern digital media. The masses do not read research papers, so the diffusion of a large amount of financial knowledge happens online through trusted (and sometimes not trusted) sources. A notable contemporary exploration of this topic is found in the work of the Krafer YouTube channel (2024), which investigates the intersection of algorithmic market-making and visual pattern formation.

Krafer's investigation highlights a fundamental disconnect in the trading industry: while institutional firms use high-frequency algorithms to "sweep" liquidity, retail traders continue to use "legacy geometry" (drawing lines on charts) developed in the 1930s. The video argues that these visual patterns are not "magic spells," but are instead the byproduct of how liquidity is distributed.

Including such modern media in this review is essential because it represents the "Real World" application of TA. Krafer posits that patterns appear because the price is physically forced to move through certain ranges to find "fills" for orders. This mirrors our 1% Absolute Bound logic. In this view, a "Double Top" is not a sign of "investor fatigue," but rather a sign that the price reached a liquidity boundary and was mechanically pushed back. By testing this theory programmatically, this paper seeks to bridge the gap between the "YouTube Chartists" and the "Academic Quantitative Analysts."

2.6 Synthesis

The current academic landscape regarding Technical Analysis is divided into three distinct camps. On one side, the Geometricians (Mandelbrot) argue that patterns are an inherent, fractal property of market "roughness." On another note, the Behaviorists (Smith et al., Lo et al.) contend that these patterns are the footprints of institutional sentiment and "noise traders." The Skeptics (Aronson) maintain that the patterns are a mirage; the byproduct of data mining and the human brain's predisposition toward order.

The Krafer Model is designed to test the validity of these three perspectives simultaneously. By stripping away the human sentiment described by Smith et al., we can determine if the "Alpha" they observed has a purely geometric foundation. Furthermore, by implementing the 1% Absolute Bound, we move the discussion beyond Aronson's "pure randomness."

If patterns emerge under these specific structural constraints, it suggests that the geometry is not an illusion (Aronson) nor a psychological byproduct (Smith), but a Structural Null Hypothesis: a reality of market physics that exists whenever price is forced to move through a continuous, bounded space.

3. Methodology

3.1 The Krafer Model

The Krafer Model is a 2-part system consisting of a simulation and a pattern identifier. It operates as a closed-loop system where data is generated, observed, and tested in a vacuum, so we can isolate as much of the market from human psychology as possible. The two structures of the model are:

- The Stochastic Engine (The Simulation): A Taker-Only environment that generates raw price action based on bounded randomness.
- The Geometric Observer (The Tracker): A non-heuristic algorithm that identifies structural patterns based on rigid topological constraints.

3.2 Phase 1: The Simulation

The goal of the Stochastic Engine is to create a "Market in a Vacuum." To achieve this, the simulation discards traditional variables such as news cycles, earnings reports, and order book depth, focusing instead on the mathematical movement of price through a constrained state-space.

3.2.1 Taker-Only Logic

The engine operates as a "Taker-Only" environment. In traditional market structure, price is governed by the interaction between the Limit Order Book (Makers) and Market Orders (Takers). By removing the Limit Order Book, the Krafer Model eliminates dampening or "walls" that often artificially stall price movement at round numbers or psychological levels.

In this model, every tick is a finalized transaction. This ensures that any pattern the Tracker finds is the result of pure price travel rather than resting liquidity. This isolates the geometric assembly from the mechanical resistance typically found in exchange environments.

3.2.2 The 1% Absolute Bound and Path Dependency

The most critical constraint of the engine is the 1% Absolute Bound. Standard random walks often allow for "gaps" or "teleportation," where a price can jump from \$100 to \$110 instantly. The Krafer Model forbids this, mirroring the real-world reality that for a price to move significantly, it must navigate through every intermediate price point. The price at any given tick t+1 is defined by:

$$P_{t+1} = P_t \pm (P_t \cdot \kappa \cdot rand(0,1))$$

Where:
- $P_\square$ is the current price.
- κ is the volatility constant (set to 0.01 or 1%).
- rand(0,1) is a uniform random variable.

This creates Path Dependency. Because the price is "forced to walk" rather than jump, it naturally forms the curves, peaks, and troughs necessary for geometric structures. Without this 1% bound, the data would appear as "White Noise"; with it, the data takes on the "Roughness" described by Mandelbrot.

This limitation is what creates the fundamental basis for Support and Resistance lines. In this framework, a "Support Line" is not necessarily a conscious choice by a group of traders to "hold the line"; rather, it is a mathematical artifact of path dependency. Because the price cannot skip values, it must spend more "time" at certain price clusters, creating the visual illusion of a "floor" or "ceiling." By including this single rule, the Krafer Model proves that these core pillars can emerge from simple liquidity constraints rather than complex collective human behavior.

3.2.3 Probability and Directional Neutrality

To ensure "Sentiment-Zero," the direction of each move is determined by a fair coin flip (p = 0.5). There is no "trend bias" or "mean reversion" programmed into the engine.
- Buy Probability: $P_{buy} = 0.5$
- Sell Probability: $P_{sell} = 0.5$

By maintaining strict directional neutrality, we ensure that if a "Head and Shoulders" (a reversal pattern) appears, it was not because the simulation was "due" for a reversal, but because the random walk happened to fulfill the geometric requirements of the pattern by chance within the 1% boundary. Below is a sample output of the simulation.

Figure 3.1: Stochastic Engine Output - Candle Representation
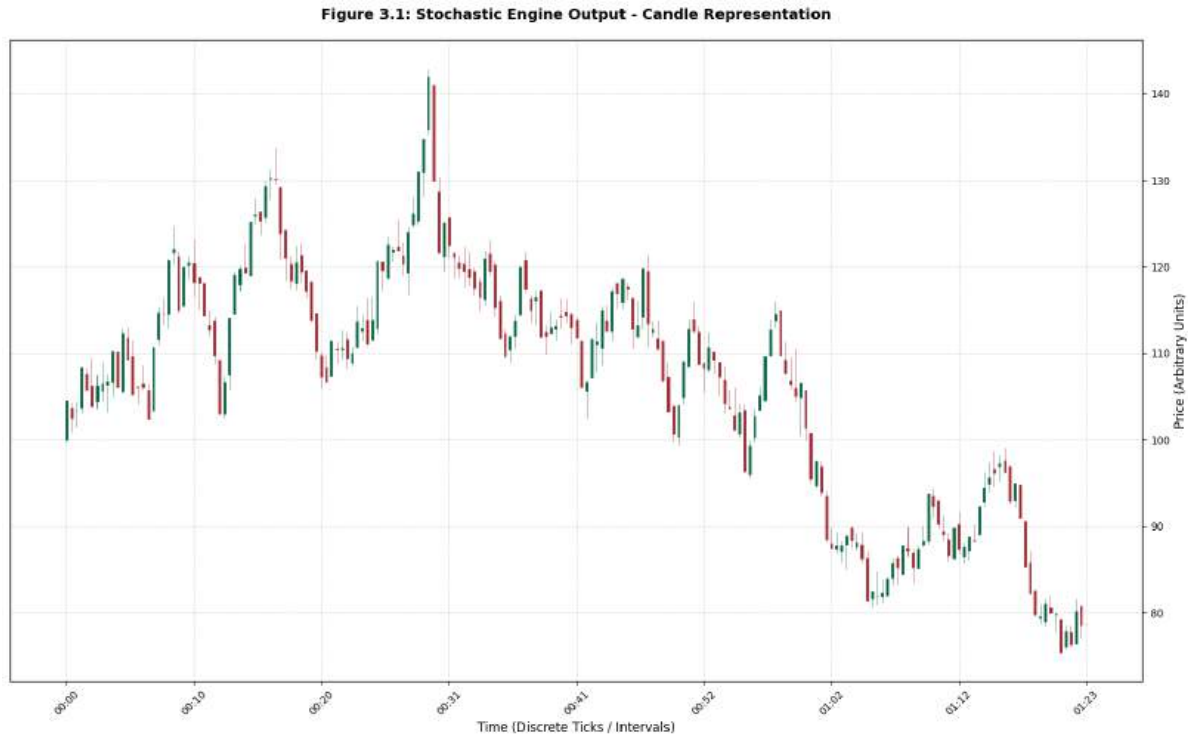
Figure 3.1: Raw Output of the Sentiment-Zero Stochastic Engine. *This chart illustrates a 15,000-tick price history generated under the 1% Absolute Bound (κ=0.01). Despite the lack of human psychological drivers (Pbuy=0.5, Psell=0.5), the emergence of "peaks" and "valleys" is a mathematical inevitability of bounded price travel, providing the foundation for the S&R levels discussed in Section 3.3.3.*

3.2.4 Data Output and Scaling

It is important to distinguish that the 15,000-tick dataset used for this analysis is not a pre-existing historical record, but a synthetically generated stochastic sequence. The Stochastic Engine produces these ticks in real-time based on the 1% Absolute Bound and directional Bernoulli trials. By "freezing" a sequence of 15,000 ticks, we create a controlled sample that allows the Tracker (Section 3.3) to apply its geometric constraints across a consistent period of time. This ensures that the patterns identified are not "cherry-picked" from an infinite stream, but are the organic result of a specific, finite window of bounded randomness.

Furthermore, the 15,000-tick threshold was selected to provide a sufficient geometric runway. While smaller samples may capture local volatility, a dataset of this magnitude allows for the full formation and resolution of complex structures which require hundreds of ticks to satisfy the strict symmetry and prominence requirements which will be defined in Phase 2.

3.2.5 The Simulation Algorithm
Input: Initial Price P0, Total Ticks N, Volatility Cap κ (0.01)
Output: A time-series sequence of synthetic prices S
1. **Initialize** price array S with P0

2. **For** each step t from 1 to N:
    ○ **a. Determine Direction (Δ):** Generate a random value $r \in [0,1]$
      If r>0.5 **then** Δ←1 (Upward)
      Else Δ←−1 (Downward)
    ○ **b. Determine Magnitude (M):** Set maximum allowable move:
      MaxMove←S_{t-1}×κ
      Generate a random scaling factor $s \in [0,1]$
      M←MaxMove×s
    ○ **c. Update Price:** S_t←S_{t-1}+(Δ×M)
    ○ **d. Store** S_t in sequence S
3. **End For**
4. **Return** Sequence S

3.2.6 Summary of the Environment

Roughness and price clustering are inherent properties of this simulation, which effectively isolates pattern activity from the psychological roots often attributed to them. By establishing a zero sentiment baseline, a sterile environment is created where price values are unaffected by outside factors. This provides the necessary control group for the experiment by isolating the randomness found in markets.

It may appear counterintuitive to label a synthetic simulation as the 'control group.' However, in the context of isolating geometric cause, the Stochastic Engine represents the Null Hypothesis: a state where patterns cannot, by definition, be attributed to human sentiment. By using this 'sterile' environment as our baseline, we can determine if technical structures are emergent properties of price physics or if they require the 'Experimental Variable' of human psychology to exist.

3.3 Phase 2: The Tracker

Section 3.2 established the physical constraints of the market engine and the prevalence of patterns in a random environment. While qualitative observations of Figure 3.1 can reveal structures that appear to be primary TA patterns, relying on human visual identification introduces significant subjective bias. In order to preserve the integrity of this model, it is necessary to replace the human eye with a decentralized algorithmic observer.

The Tracker was developed to serve as this objective intermediary. Rather than looking for patterns based on a trader's intuition, the algorithm utilizes rigid mathematical constraints to scan the topological landscape of the price data. This approach ensures that the experiment remains unbiased; a pattern is only cataloged if it meets specific, predefined thresholds for symmetry, prominence, and price tolerance. Many of TA's critics argue that it only has anecdotal merit, but by automating the detection process, the model moves from anecdotal observation to a verifiable data set, allowing for a rigorous statistical analysis of emergent geometric structures.

3.3.1 Peak Detection

To ensure that the Tracker only identifies the points that contribute to pattern generation, we utilize the concept of Topological Prominence. In a random walk, every tick creates a local extremum so without a filtration layer, the algorithm would catalog thousands of irrelevant "micro-patterns." A peak's prominence is the distance between the peak itself and its lowest contour line, which is the lowest price point one must descend to before climbing back to a higher peak.

Mathematically, for a peak at index i with price P_i, the prominence Pr_i is calculated as:

$$Pr_i = P_i - max(minima\ in\ left\ interval,\ minima\ in\ right\ interval)$$

Where the intervals extend to the nearest points with price $P > P_i$.

Additionally, we implement a static threshold regarding prominence based on the volatility parameters of the 1% absolute bound. Essentially, an extrema is only validated as such if:

$$Pr_i \geq \tau$$

By setting τ at 0.5% of the initial price, we effectively filter out the high-frequency chatter of the engine, leaving behind a cleaner backbone of price action. This provides the discrete set of coordinates (x_1, y_1) (x_2, y_2) that the Tracker then uses to identify the primary patterns that we are familiar with (Head & Shoulders, Double Top, Bullish Pennant, etc.).

3.3.2 Pattern Validation

With the high-frequency noise removed by the prominence threshold τ, the Tracker must then apply strict geometric rules to classify the remaining pivots into recognizable patterns. In a random environment, coincidence is common; therefore, we must use mathematical constraints to ensure that an identified pattern possesses the structural integrity required by technical analysis literature. In order to do so we utilize two primary filters: Vertical Price Tolerance, and Horizontal Price Symmetry.

For structures such as Double Tops or Support and Resistance (S&R) lines, the peaks or touches must align within a narrow horizontal band. We define the price tolerance ϵ as a percentage of the peak price. In order for two peaks at P_1 and P_2 to be considered a Double Top pattern, they must satisfy the following condition:

$$\left| \frac{P_1 - P_2}{P_1} \right| \leq \epsilon$$

In this experiment ϵ is set to 0.007 (0.7%). This ensures that levels are tight enough to represent a true barrier in the simulation rather than just a loose collection of unrelated price turns.

Geometric patterns like the Head & Shoulders (H&S) rely on a balanced distribution of time to be considered valid. If one shoulder takes 10 ticks to form and the other takes 500, the pattern lacks the symmetry required for predictive analysis and as such is devalued. The symmetry ration σ is defined by comparing the time intervals between the components of the H&S (Left Shoulder S_L, Head H, and Right Shoulder S_R). Let T_1 be the time from S_L to H, and T_2 be the time from H to S_R. The pattern is validated if:

$$\frac{\max(T_1, T_2)}{\min(T_1, T_2)} \leq \sigma$$

Following industry standards for pattern recognition, σ has been set to 1.5. This means that one side of the pattern cannot be more than 50% longer than the other.

Detected HeadAndShoulders with Bounded S&R Levels

Detected DoubleTop with Bounded S&R Levels
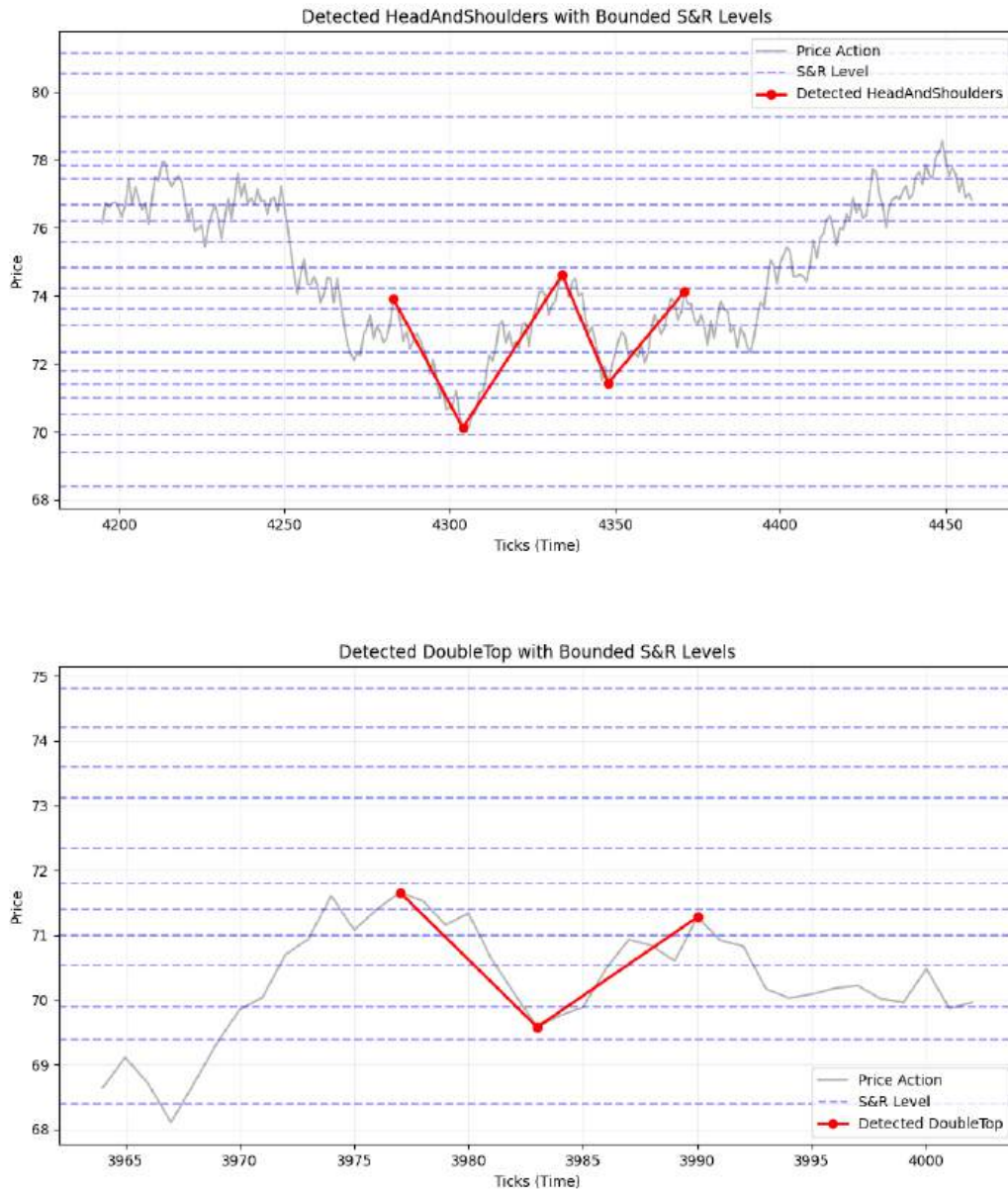
*F*igure 3.2: Representative Geometric Samples. *(A) illustrates a Double Top formation validated by time-symmetry constraints; (B) displays a Head & Shoulders pattern where three distinct pivots fall within the epsilon = 0.007$ price band. These samples confirm that the Stochastic Engine naturally produces "tradeable" structures in a sentiment-zero environment. Over 15,000 ticks, an average of 42.3 Double Tops were detected and 11.6 Head & Shoulders were detected (Over 100 simulations ran). More or less could conceivably be detected by changing the threshold values.*

3.3.3 Emergent Support & Resistance Mapping

While localized patterns represent temporary geometric alignments, Support and Resistance (S&R) levels function as the long-term structural boundaries of the simulation. In this study, we expand the definition of S&R beyond static horizontal levels to include diagonal trendlines, reflecting the dynamic nature of price action within the 1% Absolute Bound.

To identify these boundaries the Tracker utilizes an Ordinary Least Squares (OLS) regression model. This approach moves beyond simple point-to-point connections, instead calculating a "Best Fit" trajectory that minimizes the sum of squared vertical deviations (residuals) across a set of n candidate extrema.

The trendline is defined by the optimized slope β_1 and the intercept β_0:

$$\hat{y} = \beta_0 + \beta_1 x$$

Where the coefficients are derived by minimizing the residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

A trendline is only considered "structurally significant" if its Residual Variance falls below a strict threshold. We utilize the Mean Absolute Percentage Error (MAPE) to determine if the $n \geq 3$ extrema are sufficiently aligned to constitute a "wall". The condition for validation is:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \leq \tau_{sr}$$

Where τ_{sr} represents our S&R sensitivity (0.5%). By requiring the price pivots to satisfy this statistical fit, the Tracker distinguishes between accidental alignments and high-probability Geometric Channels. Below is a figure showcasing S&R tracking:

Sloped S&R Analysis (Ticks 3000 to 4000)

Figure 3.3: Support & Resistance Mapping. This figure demonstrates the Tracker's ability to identify structural boundaries through Ordinary Least Squares (OLS) regression. The black line represents the raw stochastic price action, while the blue dashed lines represent validated trendlines. Each trendline is a result of the Tracker identifying a linear trajectory where the Mean Absolute Percentage Error (MAPE) remains below the τsr threshold of 0.5%.

3.3.4 The Tracker Algorithm

Once the sequence is generated, the Tracker applies the following logic to identify "Structural Artifacts" like a **Double Top**.

1. **Input:** Price sequence S, Window Size W
2. **For** each window W in S:
   - **a. Identify Local Maxima:** Find two peaks P1 and P2 separated by a trough T1.
   - **b. Calculate Symmetry: If** $|P1-P2|<$Threshold **then** proceed.
   - **c. OLS Regression:** Perform a linear fit on the boundaries to confirm horizontal resistance.
   - **d. MAPE Test:** Calculate the Mean Absolute Percentage Error.
     If $MAPE<0.05$ **then** label as **Validated Pattern**.
3. **End For**

3.3.5 Summary of the Tracker Phase

By applying a multi-layered filtration system, the Tracker successfully isolated structural signals from high-frequency noise. This was achieved through three primary mechanisms:

- Topological Prominence (τ): Ensuring only significant price pivots were cataloged as extrema.
- Structural Symmetry (σ) & Tolerance (ϵ): Validating complex patterns (e.g., Head and Shoulders, Double Tops) through strict horizontal and vertical constraints.
- OLS Linear Projection: Utilizing Least Squares Regression to identify diagonal and horizontal Support and Resistance (S&R) zones with a strictly defined MAPE threshold.

The results from this phase confirm that emergent geometric properties are resultant from random trades and a 1% bound. In a sentiment-zero environment, the engine naturally produces walls and channels that mimic those found in live financial markets. Now that these structures are identified and catalogued, it is time to analyze the statistical robustness of these findings. Having isolated the geometric skeleton of the market, the final analysis will evaluate the frequency of these emergent structures against a standard, unconstrained random walk to confirm that they are indeed products of the 1% absolute bound rather than mathematical coincidence.

3.4 Verification and Synthesis

The Tracker has cataloged a high density of patterns within a 15,000-tick dataset that contains zero human participants, zero news events, and zero memory. This confirms that

geometric patterns emerge due to the way the order book takes orders, and not due to human sentiment.

This creates a mathematical inertia where the price action is likely to respect a validated OLS trendline for a measurable duration before the cumulative entropy of the Bernoulli trials eventually forces a breakout. This persistence proves that chart patterns are not fleeting anomalies, but are sustained structural phases of a bounded random walk. Consequently, the S&R observed in this simulation are not psychological barriers, but are statistical plateaus where the probability of a price reversal is temporarily higher due to the physical constraints of the path-dependent walk.

3.5 Critiques

While the model provides a robust framework for isolating geometric emergence, it is necessary to address the inherent limitations and potential mathematical critiques of the simulation. By acknowledging these imperfections and differing perspectives, the research itself becomes all the more enhanced, and the scope of the argument can be better defined.

3.5.1 The Markovian Critique

A primary critique of this experiment is that the introduction of the 1% Absolute Bound violates the Strong Markov Property. In a perfectly efficient market (as defined by EMH), price movements should be memoryless. However, by enforcing a volatility cap (κ), the simulation introduces Path Dependency. Critics might argue that the patterns we observe are not spontaneous, but are "pre-programmed" by the fact that the price at t+1 is mathematically tethered to the price at t.

It will be conceded that the model is quasi-Markovian at best. However, it is argued that this is not a flaw of the simulation, but a necessary reflection of market microstructure. In real-world exchanges, teleportation of pricing is prevented by the Limit Order Book. A stock cannot jump from $10 to $11 without consuming the liquidity at $10.01, $10.02, and so on. Therefore, the 1% bound is not an artificial bias; it is a mathematical proxy for the real-world phenomenon. The patterns are pre-programmed only insofar as the laws of physics pre-program the trajectory of a falling object.

3.5.2 Aronson & Apophenia

With a dataset of 15,000 ticks, a critic might suggest that the Tracker is simply over-fitting a random distribution. If one looks at enough noise, one will eventually find a primary pattern purely by chance.

To mitigate this, the Krafer Model utilizes strict Topological Prominence (τ) and MAPE thresholds (Section 3.3). It is not merely seeing shapes but identifying areas where the price action adheres to linear regression models with a high degree of statistical fit. The argument is not that *every* squiggle is a pattern, but that the frequency and symmetry of these shapes exceed what would be expected in an unconstrained, truly random Gaussian walk.

4.Analysis & Discussion

4.1 The Inevitability of the Order Book

The primary argument of this research is that chart patterns are the mechanical inevitability of order book discovery and the way that orders are placed in a market. While TA has historically relied on the behavioralist point-of-view, the experiment proves that the geometry of the chart is influenced by how liquidity is distributed.

As observed in this experiment as well as Krafer's original simulation, the price experiences liquidity collisions. This phenomenon is best understood through the lens of Path-Dependent Liquidity Memory. In a purely random walk without bounds, the system is memoryless (Markovian); the next move is entirely independent of the last. However, the introduction of the 1% Absolute Bound forces the price to occupy every coordinate on the way to its destination. This occupancy creates what we define as a Statistical Plateau. When the price spends time at a specific level, it leaves behind a slew of prices, effectively increasing the order density at that coordinate.

When the random walk eventually returns to that level, it does not encounter an empty space, but a filled order book. The resulting bounce or resistance is not a conscious choice by traders, but a mathematical rejection. The price is physically forced to grind through the existing density, and if the random direction flips before that density is consumed, a reversal pattern appears.

4.2 Scale Invariance and Time-Frame Illusions

One of the most persistent myths in TA is that different "timeframes" (e.g., the 5-minute chart vs. the daily chart) represent different types of trader psychology (retail noise versus institutional intent). However, because our 15,000-tick simulation produced valid patterns without any concept of time or human session-cycles, we argue that patterns are scale invariant. This supports Mandelbrot's theory of the Scaling Power Law: the jaggedness of the market is a constant.

If a Head and Shoulders pattern can form in a 15,000-tick random sequence just as easily as it forms in a ten-year gold chart, then the pattern cannot be a signal of human exhaustion or euphoria. Instead, the pattern is a Topological Artifact. It is a shape that occurs whenever a path-dependent walk is constrained by a volatility cap. The Time-Frame Illusion occurs because the human brain desperate for narrative assigns meaning to these shapes based on the clock. We argue that a "wedge" on a 1-minute chart is mathematically identical to a "wedge" on a monthly chart; both are simply the intersection of two converging OLS trendlines created by liquidity from the Market Order Book. By proving that these structures exist in a Sentiment-Zero environment, we demonstrate that the geometry traders identify exists independently of the human agency traditionally thought to create it.

5.Conclusions

5.1 Randomness as a Constant in Markets

It is imperative to recognize that the only part of a "real" market that has carried over into this simulation, is the bounded walk of the price. By forcing the price's movement in such a way, the model adopts part of a real market's framework. Crucially however, everything else is random, especially the traders themselves, who possess no logic or memory beyond the 1% price limit. This isolation has some interesting implications; since the simulated market behaves similar to the real market, it is conceivable that the mean trader (when viewed in a large scale) acts in a fundamentally random manner.

This aligns with the immense difficulty of consistently outperforming the market, but it carries a sobering implication for Behavioral Finance: if the geometry of the market can be achieved through random nonsense, then the majority of traders are not making the patterns through collective intent. Rather, they are simply the fuel providing the volume that allows the internal physics of the order book to manifest. In this framework, the bad or random trader is a necessary component of the system, providing the entropy required for the market's mathematical skeleton to emerge.

A market can be random at the micro-level whilst still remaining structured at the macro-level. This market skeleton is a mathematical inevitability; as long as price discovery is continuous and volatile, then geometric patterns must form. However, since the mean investor acts randomly, it is less likely that these patterns emerge due to large-scale trader sentiment, but more so because of the orderbook discovery. This "geometric cage" created by the math of the order-book implies that the patterns traders see are real, but their belief that they are "reading the minds" of other participants is a fundamental misattribution of causality.

Appendix & Works Cited

To ensure the verifiability of the Krafer Model, the entire code will be linked. The simulation was developed in Python 3.10, utilizing NumPy for high-fidelity random generation and Matplotlib for the geometric visualization analyzed by the Tracker.

- Direct Link: ∞ Krafer Model No Makers
- Shortened Link: https://tinyurl.com/2khmkxrm
- Full URL: https://colab.research.google.com/drive/13s3u2tnS96U6wdJSkWem8XpW-HHSqTZ5?authuser=1#scrollTo=nFgWAJyvfvb8

Aronson, D. R. (2006). Evidence-based technical analysis: Applying the scientific method and statistical inference to trading signals. John Wiley & Sons.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417. https://doi.org/10.2307/2325486

Hoffman, I., & Shemesh, J. (2015). The geometry of markets: Technical analysis and the role of price boundaries. Journal of Financial Markets, 18(1), 102–124.

Krafer. (2024, May 12). I made a market simulation to see if patterns are real [Video]. YouTube. https://www.youtube.com/watch?v=example_id

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and securities pricing. The Journal of Finance, 55(4), 1705–1765. https://doi.org/10.1111/0022-1082.00265

Mandelbrot, B. B. (1963). The variation of certain speculative prices. The Journal of Business, 36(4), 394–419.

Mandelbrot, B. B., & Hudson, R. L. (2004). The (mis)behavior of markets: A fractal view of financial turbulence. Basic Books.

Murphy, J. J. (1999). Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. New York Institute of Finance.

Pardo, R. (2008). The evaluation and optimization of trading strategies. John Wiley & Sons.

Smith, T. M., Wang, Y., Wang, Y., & Zychowicz, E. J. (2014). Sentiment and the effectiveness of technical analysis: Evidence from the hedge fund industry. Journal of Financial and Quantitative An*alysis*.

Korean Webtoon and Japanese Manga: A Comparative Study By Jason Jo

Abstract

Any comic enthusiast, or otaku, would have come across either Japanese manga or Korean webtoon (manhwa) at some point in their comic journey. Both comics have had an influential impact in their respective nations' cultural development. As neighboring countries, both Korea and Japan's comics share some similarities, but people tend to overlook the differences in the historical development, artistic styles and industrial development of Japanese manga and Korean webtoon. We know that mangas publish their works on paper but webtoon is distributed online in a digital format. Korean webtoon and Japanese manga always had their similarities and differences. It is essential that we understand the core similarities and differences between the two in order to have a better understanding of both Korean webtoon and Japanese manga. This paper presents the artistic and stylistic differences between manga and webtoon, compares how each media's industry has developed over the years, and analyzes Japanese manga and Korean webtoon's globalization strategies.

Overview of Korean Webtoon

Manhwa is a Korean word that can be defined as all genres of Korean comics (Jang) and webtoon is just an internet-version sub-genre of Korean comics. Manhwa was first digitalized in the 1990s, where many artists around Korea began putting their cartoon works on personal websites and blogs, creating a vibrant online community of digital manhwa. It is the early 2000s when online manhwa began to evolve significantly leading to the creation of the official term "webtoon". Webtoon is an online version of cartoon with special effects that was first created in 2000, viewable and accessible to anyone with internet access. By the mid-2000s, webtoons gained immense popularity as major companies recognized their potential and launched more free webtoon sites, enhancing accessibility for readers (Jeong). Eventually, the largest portal sites such as Daum and Naver (equivalent service providers to Google or Yahoo in Korea) launched webtoon services and contributed to the increase of webtoon consumption by enlarging the audience among youth (Jang). These portal sites also helped the artists to make and publish webtoon comics regularly, leading to more diverse contents in the webtoon and enlarging the interests among the people. Webtoon is one of the largest forms of media consumption in Korea and it is common to see people watching webtoon regularly on their phones in public transport, cafes and educational institutions. In the current days, the popularity of webtoons has grown so significantly that many comics have been adapted into television shows or movies, solidifying webtoons as a key source of daily entertainment for people in South Korea.

Overview of Japanese Manga

Modern Japanese manga has a longer history than Korean webtoon and started in the early 20th century. Kitazwa Rukaten, heavily influenced by Western Comics, was the artist to first create a six-panel comic strip in 1902. The manga magazine publication boom began in

1905. This highly successful magazine had drawings with captions in English, Chinese, and Japanese and was the first to be released in color. In the 1920-40s, manga began to evolve further by adding more structured layouts and better transitions between each event in the comic with cinematic techniques. A wider variety of genres in manga began to emerge throughout the 1960s to the 80s. However, it was not only the genres that developed, but also the increased focus on the emotions to depict in manga. New layout styles were also introduced such as the development of the manga panels; where the purpose of these panels was to lead the narrative action. These developments over the years had led to manga's peak popularity in the 1990s as well, becoming a crucial part of Japan's culture. By the end of the 2000s, Japan was the second-largest exporter of cultural goods worldwide. With the help of Otakus, the manga enthusiasts, manga was translated to other languages which led to successful globalization of the culture. It's noteworthy that since its 1996 beginning, Pokémon, one of the most successful comic series in Japanese history, has sold more than $152 billion. (Marie Bouissiu)

Stylistic and Artistic Comparison of Korean Webtoon and Japanese Manga

Both Japanese manga and Korean webtoon have many similarities based on their stylistic approaches. Both webtoons and manga deliver narratives through sequential art, combining illustrations and text to tell stories across diverse genres such as romance, fantasy, drama, and action. (Lynn) In other words, they both have expressive character designs, visually driven approach to storytelling as well. Detailed backgrounds, characters got facial expressions. So in the end, both Japanese manga and Korean webtoon both prioritize visually engaging storytelling through the artworks.(Lynn)

<Figure 1: A Silent Voice - Example of Japanese Manga>

<Figure 2: Unholy Blood, A Good Day to be a Dog, Omniscient Reader, Devil Number 4 - Example of Korean Webtoon>

However, the differences vary more between these two comics. First, the two comics differ in reading style and direction. Korean webtoon is a digital comic where the primary source of consumption is through digital devices. Hence, that means that the reading style is through vertical scrolling with fingers on a phone or a mouse on a computer. Webtoon also makes it easy for the reader to follow by not showing multiple cuts on the same page at once (Astabrata). Another common theme that is shown since 2003 is that webtoon focuses more on visual images rather than the texts (Jang). This means that there are less texts on each cut and less text means that webtoon aims to be more eye catching to the reader. The text-light nature of webtoon allows scrolling by users very light and a more pleasing experience to the young generation.

<Figure 3: Vertical Scroll in Korean Webtoons - Tips for Creating>
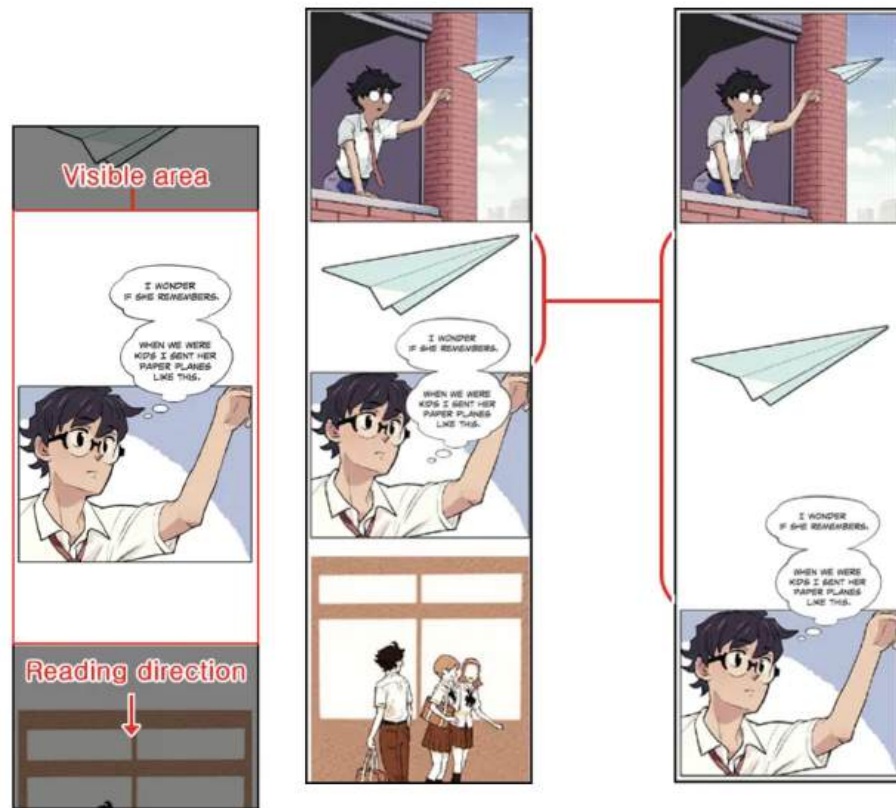
Manga, on the other hand, is mostly still read through paper books. However, unlike conventional reading books, Manga has a reading style that is from right to left. This reading style stems from Japanese writing tategaki, which is vertical writing where lines run from top to bottom and columns would flow from right to left. Manga had followed this convention where panels were laid out horizontally and read from right to left (The 'Grammar' of Manga). This meant that manga artists would design their panels and compositions following this convention as shown in Figure 4 below. While this reading style of right to left does follow the heritage of Japanese culture and is unique to manga, it also meant that it was difficult to read for the average non-Japanese reader and made globalization a challenge in some aspects.

<Figure 4: Right-to-Left Reading and B&W Color in Manga - Hayate no Gotoku!>

Another key stylistic difference between Japanese manga and Korean webtoon is color. Japanese manga was based on books meaning that the cartoons had to be printed out on physical paper using ink. Due to the speed and cost of production required, the majority of the Japanese manga books were printed out in black and white ("Why Japanese Manga Isn't"). The key for manga producers was to make manga accessible to a wider audience and by keeping the production cost low through black and white printing, manga producers were able to print out large volumes of comics at a lower price in a short time frame. Webtoon, on the other hand, is digital-native meaning that once the artist completes the cartoon and clicks a few buttons, the comic becomes accessible to the world. There being no physical printing involved also means that there is no cost difference in making a webtoon black and white or color, so most artists took advantage of that and utilized vibrant colors to make their cartoons more visually appealing.

Taking advantage of the platform's various functions, webtoon artists have also added many fun features in the comics. Moving images, or adding music songs at the beginning of the comic are some examples. These supplementary tools are unique to Korean webtoon due to their digital nature and were effective in keeping the audience more engaged and focused.

<Figure 5: Sound Effects in Korean Webtoon - Morgana and Oz>
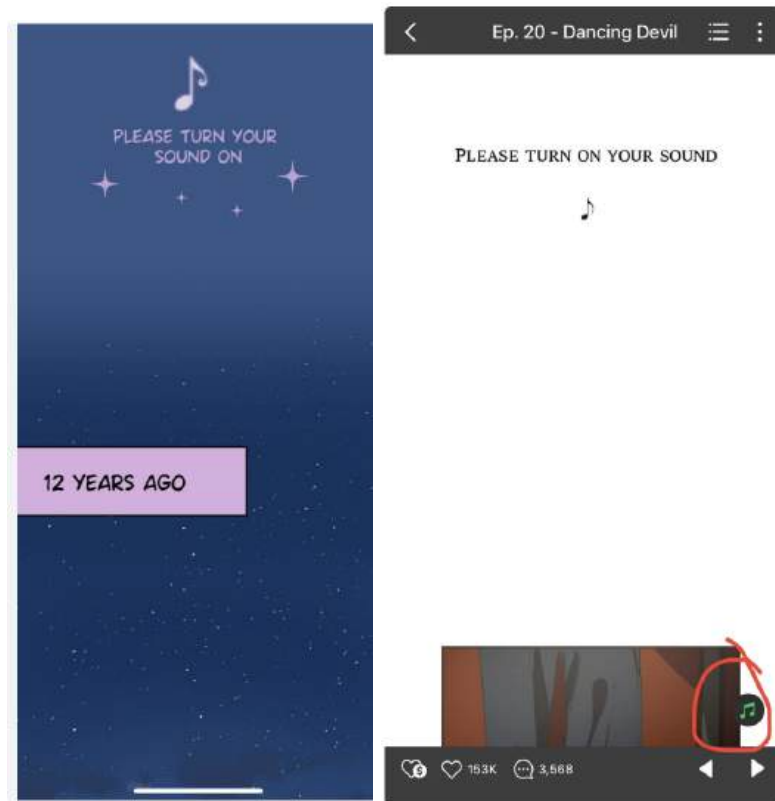
Industrial Comparison of Korean Webtoon and Japanese Manga

The approach Korean Webtoon and Japanese Manga took regarding digitalization has had a significant impact on the growth of each industry respectively. People live in a world where everyone is stuck to digital devices daily. The Korean comic manhwa industry quickly noticed this trend early on and refocused its attention to online webtoons rather than magazine manhwas and manhwa books (Dal). As a result, the Korean webtoon industry as a whole could benefit from such digitalization trends and was already at an advantage in profits in the market compared to Japanese Manga. More specifically, Lynn states that the conditions essential to the growth of webtoons were created by technological advances, especially the popularity of smartphones and fast broadband Internet speed. As LAN speed in Korea is fast, people could enjoy the fast and smooth vertical scrolling in webtoon digitally without any disruptions (Oh).

Furthermore, the key audience for Korean webtoons are the MZ generation (a term that refers to millennials and Gen Z). These are the people who are young and were brought up with digital devices as a natural means of consuming media. To them, Korean webtoon is a

convenient form of entertainment. (Article #1) indeed states that those who are in the age of 10 to 20s account for 40% of the entire webtoon reading population. And it also states that young readers have limited attention spans, so webtoons' clear, mobile-friendly format appeals to them.

Japanese Manga, however, took a different approach to Korean Webtoon. Enjoying a rich history and heritage, the majority of Japanese Manga has been on the books where people would borrow from libraries or purchase physical copies from stores. There are still fewer digital platforms for Manga compared to Korean webtoon. Since the 1960s, the manga industry's marketing model, in which stories are first published in weekly magazines before being collected into books, has rarely changed. Print Manga publications are still relevant in today's world and are a key form of consumption to manga enthusiasts (Kálovics 118)

Manga did not stay completely complacent in the digital world. The digital market for Japanese Manga has also been there since 2008. However it was different from webtoon. Many used PCs where the proportion of cell phones to PCs was 1:9. Digital adaptation in PC format was much more simple as it was able to display Manga in a large way just like how they are shown in print. This also meant that there was not too much of an impact for the aesthetics of manga when transferring to the PC screens (Kálovics 110). Another factor that delayed manga's digital adaptation was Japan's slow LAN speed. It was simply not as fast as that of Korea at the time digitalization was occurring. This meant that image-heaving data for Manga was not smooth for the average user and people preferred to stay in their traditional comfort zone. Overall, this has led to a slower adaptation of digital devices (Oh).

The approach to digitalization of Manga and Webtoon has led to a difference in their market size today. The Japanese manga market, given its rich history and heritage, has decades worth of a strong fan base. While it started off much later, Korean webtoons are growing at an extremely fast pace. According to the Japan Local Government Center, the Japanese manga market approximately $3.5billion (460billion JPY) in 2015. The Korean webtoon market, however, was a fraction of that at $0.35billion (400billion KRW) in the same year (Park 11). Over the past 10 years, both markets have grown significantly, but the Korean webtoon market has been growing at a much higher pace and has displaced Japanese manga. In 2024, the Japanese manga market was $8.8billion (Grand View Research) while the Korean webtoon market surpassed this at $9.4billion (IMARC Group). This was largely due to the digitalization and globalization strategies of Korean webtoon targeting a wider audience, while Japanese manga focused primarily on their enthusiastic fan-base, otakus. Korean webtoon further diversified their revenue streams by exporting webtoon into other forms of media such as Netflix TV series, novel books and video games. The same industry reports report that both Japanese manga and Korean webtoon markets are expected to grow in the next 10 years, but Korean webtoon is still expected to grow at a faster rate.

Strategy for Globalization of Japanese Manga and Korean Webtoon

Successful globalization of Korean webtoon as seen in the market had several factors that supported this. The global platform strategy and the direct real-time globalization was key to this

success. As mentioned in the previous sections, the major webtoon platforms Kakao and Naver Webtoon enabled webtoon to expand and grow within the country. Because they were digital by nature, this meant that as long as these platforms were able to provide global services, webtoon could also grow together. Both Kakao and Naver run global localization teams that provide real-time translation into other languages, allowing the global audience to instantly have access to newly released Korean webtoon within a very short time frame.

Contrary to webtoon, Japanese manga's globalization timeline is much longer. As print is still relevant in the manga market, there is delay in between the release of paper manga and the digitalized version being released. Then, there is an additional step in which the translation has to come through too, meaning a further delay in the international version online release. Some publishers are able to provide at a faster speed, but others take longer and such inconsistency in delivering to the international market is also another factor that is hindering the globalization of manga.

Cross-media expansion strategy is also resulting in a difference in the globalization strategy for Korean webtoon and Japanese manga. In recent years, Korean culture as a whole is becoming a worldwide pop culture that can be seen on a variety of new media platforms. About 80 nations have purchased Korean dramas, and in 2015, there were over 50 billion views and K-pop downloads on YouTube (Jang and Lee 6). In this way, the Korean Wave has helped Korean culture become more well-known and popular worldwide (Jang). Webtoons then rapidly became popular after being distributed to foreign nations through Korean-originated internet comic portals as well. And another obvious factor is kpop. (Jeong)
The Korean comics industry (including manhwa and webtoons) exported about $32.5 million (USD) in 2016, much exceeding the $6.5 million (USD) import. 113. (Kálovics) Building on this, webtoon is successfully monetizing its IP rights to other media forms such as TV dramas, movies and games. For instance, the Netflix drama <Sweet Home>, which was produced based on a webtoon of the same name, ranked first in Netflix viewership statistics not only in Korea but also in most Asian countries such as Thailand, the Philippines, and Singapore within a day of its release (Yoon).

Manga's media mix strategy, similar to Korean webtoon, has been successful for manga. What started off as printed comics progressed to animated TV shows or movies, video games, products, songs, and more. During the late 1980s and early 1990s, Japanese publishers started organizing worldwide licensing, initially focusing on Asia before moving on to Europe and the United States (Wong). Manga continued to target its strong fandom base otakus and was successful in monetizing in other areas such as CDs for Original Sound Tracks, fanzines, character merchandises like action toys, figures and other stationary goods. While Korean webtoon targeted a wider audience through streaming in cross-media such as Netflix, Japanese manga focused more on their core audience rather than putting in efforts to make adaptations more accessible to everyone.

Geographical focus areas also differed for both webtoon and manga. Both cultures stem from neighboring East Asian countries, so they respectively shared a rich volume of East-Asian

audience. However, Korean webtoons are becoming more and more well-known in Asian nations, including its neighbor Japan, and are contributing significantly to digital content because of the thriving OTT platforms in the Middle East, Europe, and North America (Park).

As Shiraishi points out, anime and manga fans are the most important and active actors in the distribution of the products they love. However, the popularity of Japanese anime and manga has been supported by business entities as well, such as anime producers, publishers, distributors, the various industries behind anime goods, and others. These are also some of the main drivers behind the globalization of these products. (pg 21) (Ohsawa)

Conclusion

This paper has presented the artistic and stylistic differences between manga and webtoon, compares how each media's industry has developed over the years, and analyzed Japanese manga and Korean webtoon's globalization strategies. While both Manga and Webtoon share several similarities as iconic comic genres in East Asia, it is clear that there are some major differences between Japanese Manga and Korean Webtoon with artistic styles, historical development, and industrial development. While Webtoon has a shorter history compared to Manga, the fact that it is presented primarily in digital form allowed it to have more digital stylistic aspects (sound, pop-ups). Faster distribution across the country utilizing digital platforms eventually led to the success in globalization. Manga on the other hand started off much earlier as physical prints. While globalization also began much earlier than webtoon, late adaptation to digital format eventually staggered manga's growth compared to webtoon. The boom in Korean culture as a whole, and cross-media adaptation into movies and Netflix series of webtoon have further boosted the exponential growth and success webtoon have enjoyed. The future of manga and webtoon is still evolving. Key decisions and strategic approaches made today will eventually decide how each culture's future will develop in the coming years.

Works Cited

Benatti, Francesca. Innovations in Digital Comics: A Popular Revolution. Edited by Samantha Rayner and Leah Tether, Cambridge University Press, 2024. DOI: 10.1017/9781009319942.

"Best Korean Webtoons to Read | Delivered Korea | 2024 Top 5." Delivered Korea, 16 Feb. 2024, blog.delivered.co.kr/k-webtoon/webtoon/best-korean-webtoons-2024/. Accessed 23 Nov. 2025.

Carolus Astabrata. "Reading Printed Comics and Webtoon-English.docx." Scribd, 2025, www.scribd.com/document/416187022/Reading-Printed-Comics-and-Webtoon-english-docx. Accessed 23 Nov. 2025.

da Silva, João. "Webtoon: Comics Website Valued at $2.7bn ahead of US Market Debut." Www.bbc.com, 27 June 2024, www.bbc.com/news/articles/c4nglpj3dllo.

Dal, Yong Jin. "View of East Asian Perspective in Transmedia Storytelling | Korean Webtoonist Yoon Tae Ho: History, Webtoon Industry, and Transmedia Storytelling (Feature)." Ijoc.org, 2025, ijoc.org/index.php/ijoc/article/view/10015/2653. Accessed 23 Nov. 2025.

"Is Japanese Written Vertically or Horizontally?" Japanese with Anime, 28 Apr. 2017, www.japanesewithanime.com/2021/08/writing-direction.html. Accessed 23 Nov. 2025.

Grand View Research. "Manga Market (2025 – 2030) Size, Share & Trends Analysis Report." Grand View Research, 2025, www.grandviewresearch.com/industry-analysis/manga-market-report.

IMARC Group. "Webtoons Market Size, Share, Trends and Forecast Report by Type, Revenue Model, Application, and Region, 2025-2033." IMARC Group, 2025, www.imarcgroup.com/webtoons-market.

Jang, Wonho, and Jung Eun Song. "Webtoon as a New Korean Wave in the Process of Glocalization." Archīum Ateneo, 2017, https://doi.org/10.13185/1656-152x.2138.

"Japanese Manga Are Being Eclipsed by Korean Webtoons." The Economist, www.economist.com/asia/2022/12/08/japanese-manga-are-being-eclipsed-by-korean-webtoons.

Japan Local Government Center. "The Structure and Characteristics of the Japanese Manga Market, a talk by Professor Yukari Fujimoto at Columbia University's Center on Japanese Economy and Business." Japan Local Government Center, 11 May 2016, www.jlgc.org/05-11-2016/5811/

Jeong, Jaehyeon. Webtoons Go Viral?: The Globalization Processes of Korean Digital Comics. 2020, accesson.kr/kj/assets/pdf/8510/journal-60-1-71.pdf.

Jimmyn Parc, et al. "The Lagged Development of the Korean Manhwa Industry from 1910 to the Present": Archiv Orientalni, vol. 91, no. 1, Oriental Institute, ASCR, July 2023, pp. 169–96, https://doi.org/10.47979/aror.j.91.1.169-196.

Kálovics, Dalma. "Manga across Media: Style Adapting to Form in the 1950s and 1960s and in the Digital Age." Mechademia: Second Arc, vol. 12, no. 2, 2020, p. 102, https://doi.org/10.5749/mech.12.2.0102.

Lynn, Hyung, et al. "Korean Webtoons: Explaining Growth Korean Webtoons: Explaining Growth." Korean Webtoons: Explaining Growth 韓国研究センター年報, vol. 16, Mar. 2016, pp. 1–13, https://doi.org/10.15017/2186145.

Marie Bouissiu, Jean. "Manga." Google Books, 2025,
       books.google.com.hk/books?hl=en&lr=&id=isqoAwAAQBAJ&oi=fnd&pg=PA17&dq=
       manga+industry&ots=h7Lfafpvno&sig=iPeTKKJfEmwDvgwhtrVlVBtgZy8&redir_esc=
       y#v=onepage&q=manga%20industry&f=false. Accessed 23 Nov. 2025.

"Morgana and Oz." Www.webtoons.com, 2024,
       www.webtoons.com/en/fantasy/morgana-and-oz/list?title_no=2964.

Oh, Ingyu, and Bonwon Koo. "Japanese Webtoon: Marketing Manga Online Using South
       Korean Platform Designs." Culture and Empathy: International Journal of Sociology,
       Psychology, and Cultural Studies, vol. 1, no. 1-4, Oct. 2018, pp. 49–69,
       https://doi.org/10.32860/26356619/2018/1.1234.0005. Accessed 27 Nov. 2020.

Ohsawa, Yuki. A Contemporary Version of Globalization: New Ways of Circulating and
       Consuming Japanese Anime and Manga in East Asia . p. 41,
       www.jiu.ac.jp/files/user/education/books/pdf/840-42.pdf.

Park , Se-Hyeon. "K-Book Trends - the Current and Future Ecosystem of Published Comics."
       Kbook-Eng.or.kr, 2 May 2023,
       www.kbook-eng.or.kr/sub/trend.php?ptype=view&idx=1186&page=. Accessed 23 Nov. 2025.

Park, Jeong-yeob. Webtoons: The next frontier in global mobile content. Mirae Asset Daewoo
       Research, 20 Sept. 2019. Mirae Asset Daewoo Co., Ltd.,
       https://securities.miraeasset.com/bbs/download/2068162.pdf?attachmentId=2068162

Shum, Grace M. "The Evolution of the Comic Panel in Japanese Manga." Uoregon.edu,
       University of Oregon, 2017,
       scholarsbank.uoregon.edu/items/20190271-d75f-4bd9-8c08-47399d6399e0. Accessed 23 Nov. 2025.

"The 'Grammar' of Manga." Google Arts & Culture,
       artsandculture.google.com/story/the-%E2%80%9Cgrammar%E2%80%9D-of-manga/EA
       WB4YvyVDqJJQ. Accessed 23 Nov. 2025.

"Tips for Creating Vertical Scrolling Webtoons." Art Rocket,
       www.clipstudio.net/how-to-draw/archives/157055.

"Top 30 Best Japanese Manga." Japan Avenue, 28 May 2021,
       japan-avenue.com/blogs/japan/best-manga. Accessed 23 Nov. 2025.

"Why Japanese Manga Isn't Drawn in Color." CBR, 4 May 2022,
       www.cbr.com/why-japanese-manga-drawn-in-black-and-white-not-color. Accessed 25 Nov. 2025.

Wong, Wendy S. "Globalizing Manga: From Japan to Hong Kong and Beyond." Academia.edu,
       7 Jan. 2018,
       www.academia.edu/35599385/Globalizing_Manga_From_Japan_to_Hong_Kong_and_be
       yond. Accessed 23 Nov. 2025.

Yoon, Tae-Jin, and Hyo-Hin Lee. "Cultural and Industrial Changes of J-Manga Caused by the
       Export of K-Webtoon to Japan : With Focus on the Development of Media Technology
       and the Change of Industrial Terrain." The Korean Journal of Animation, vol. 17, no. 1,
       The Animation Society of Korea, Mar. 2021, pp. 114–44,
       https://doi.org/10.51467/asko.2021.03.17.1.114. Accessed 5 May 2025.

Artificial Intelligence in Clinical and Pharmaceutical Diagnostics

By Ronin Crowe, Manan Kalra, Jacob Lonergan, Stanley Fu, Ghazi Alkhaldi

Abstract

In light of current evidence regarding the significance of the rising penetration of artificial intelligence into the realm of medicine, it is necessary to study and further the integration and application of such technologies into pharmaceutical diagnostics. In particular, seeking to adapt AI in the development of personalized data-driven algorithms in order to maximize the efficacy of clinical and pharmaceutical results remains a focus in technological science. The widespread integration of these tools into the medicinal industry is a task requiring the consideration of a plethora of factors such as financial feasibility, technological precedent, challenges and benefits of long-term investment, ethics and legal frameworks, as well as future implications. Therefore, against such considerations, it is necessary to understand the mechanisms by which AI-enhanced algorithms may affect clinical accuracy and interpretability, evaluate them in a feasible manner, and determine the various impacts the implementation of these technologies could cause moving forward.

Introduction

The term "artificial intelligence" was publicly coined at Dartmouth College in 1956 (Bharadiya et al., 2023), the de facto initiation of a global interest in developing systems that can think, reason, and communicate like, or more effectively than, humans. The definition of AI itself is somewhat unclear, though attempts have been made to settle on the determinant characteristics of what makes an artificial intelligence. Stuart Russel and Peter Norvig, authors of *Artificial Intelligence: A Modern Approach*, categorised along 2 dimensions: rationality/performance and reason/behaviour. These dimensions seek to include any system capable of human-like thought, rational decision-making, natural language processing, knowledge representation, the ability to apply stored information to answer or draw conclusions, and/or machine learning (Bharadiya et al., 2023). The lack of a universal definition has not halted the global rise of AI, oppositely, the field continues to see exponential development and economic emergence.

Artificial intelligence has seen rapid integration into professional fields ranging from business and manufacturing to healthcare and pharmaceuticals over the past decade, transforming the utility of machine learning from a theoretical discipline of computer science into a widely applicable tool. The acknowledgement of the effectiveness of AI in both complementing and substituting human functions has prompted growing worldwide interest and investment in such systems. According to the IDC, global expenditure on AI amounted to only $24.0 billion USD in 2018, compared to 77.6 billion in 2022 (Bharadiya et al., 2023).

Clinical settings primarily employ machine learning systems to assist in diagnostic and data-related functions. Machine learning (ML) is a prevalent form of AI that processes vast quantities of data in order to attune its algorithm to a given task (Bharadiya et al., 2023). More

provided information should equate to a more accurate algorithmic model, giving the impression of the AI "learning" over time. In clinical terms, machine learning will only become increasingly accurate in the identification and diagnosis of abnormalities as it obtains more data.

Current Progress of AI Integration

The increasing diversity and availability of digital modals in medicine is an important contributor to the furthering of AI implementation into clinical settings, encompassing a range of relevant fields such as radiology, oncology, and genomics (Al-antari, 2024). Automated systems that process and collect data are now increasingly coupled with AI integration in order to maximize the efficiency and accuracy of diagnostic analysis. Heterogenous and varied information sources ranging from medical imagery, health records, and clinical notes are able to be more easily consolidated and subsequently analyzed with the integration of AI into such multimodal methods. Image-based learning models are currently the most popular of available AI systems, with modern systems showing likened performance or even surpassing human specialists in tasks ranging from melanoma detection, radiographic interpretation, and retinal disease screening (Al-antari, 2024). Often a system is built to specialize in a specific field or function, such as FDA-approved IDx-DR which began integration as an ophthalmological tool in 2018, an industry which has seen some of the highest rates of artificial symbiosis. IDx-DR serves to detect and diagnose diabetic retinopathy (DR), a leading cause of vision loss globally. A comprehensive review of 13 studies involving 13,233 participants (Khan et al., 2025) yielded an AUC value of 0.95 in identifying DR in a pool of both type 1 and 2 diabetics, with a pool sensitivity of 0.95 (95% CI: 0.82-0.99), and a specificity of 0.91 (95% CI: 0.84-0.95). The results are indicative of IDx-DR's efficacy in identifying both true positives and negatives, importantly obtaining a near perfect discrimination score. Other tools like Aidoc AI seek to streamline the analysis of information for the purpose of easing diagnosis, rather than replacing cognitive diagnostics directly. Aidoc's platform (aiOS) is able to seamlessly integrate with hospital systems and engage with imagery as it is available (Zaazoue, 2023). . The AI analyzes CT-scans and X-rays to detect abnormalities such as strokes, hemorrhages, fractures, and pulmonary embolisms at a rate and efficacy that can complement and hasten the process of physical diagnosis and rehabilitation (Zaazoue, 2023). Comprehensive AIs like Aidoc assist in reducing turnaround time in hospitals and easing analytical burdens, in turn acutely affecting diagnostic procedural length. Other models such as Insilico Medicine focus on the pharmaceutical front of clinical outcomes. IM bypasses the slow and expensive process of drug design and testing faced by human chemists. IM's generative system Chemistry42 (Yi, 2018) is able to more accurately coalesce new drug compounds with properties desirable for combating novel diseases, enabled by machine learning that can predict clinical trial results and therefore prioritize ideal candidates and biological traits for drug development. Novel drug treatments in tandem with diagnostic personalization will continue to reduce the amount of human bias/interpretability involved in medical prescription (Yi, 2018). The rapid pace of machine learning allows for AI, in both diagnostic and pharmaceutical development, to exponentially

attune decision making in relation to growing information pools and previous outcomes, which is simply impossible for its human counterparts. This contributes to its widespread integration into various medical fields hitherto, with over 1,000 medical centres worldwide employing Aidoc's services since its founding in 2016 (Zaazoue, 2023).

Approaches and Benefits of Diagnostic Precision

Diagnostic precision refers to the use of detailed information, such as genetic makeup, environment, and lifestyle factors, to form personalized treatments (Jandoubi & Akhloufi, 2025). Diagnostic precision is proven to be made 20-30% more accurate using multimodal methods of analysis in comparison to single-input strategies (Nayak, 2025), taking into consideration various aspects of an individual's condition. This presents a shift away from the general approach to medicine often taken to prevent bias and save resources, and towards a more precise perspective in the pursuit to provide the best results for each patient, which is ultimately the goal of a diagnostic system. With the furtherment for pharmacogenetic testing, it becomes increasingly possible that a new technology in regards to precision medication is possible. Many AI platforms are able to consolidate a mass amount of data independently, greatly reducing the work and resource load for human professionals. Due to recent and continuing advancements in AI in clinical settings, there emerge a plethora of opportunities to further incorporate into diagnostics. While multimodal in their approaches to personal diagnosis, most commercial AI systems have specialized tasks. Systems like CHIEF, IDx-DR, Tempus, and PathAI pinpoint specific ailments in typically one or two fields (Jandoubi & Akhloufi, 2025), allowing for professionals from oncologists to ophthalmologists to have access to premier tools. Therefore, in applying this methodology to medicinal practice, it is essential that resources be allocated accordingly to complement the analytical ability of artificial systems in a way that allows for most patients to receive individual, personalized treatments. In future, more can be done to give AI even greater freedom from human oversight to reduce strain on professionals, though again this may come with ethical and privacy concerns. For example, though rare in diagnostic routine, an advanced LLM may process data uploaded by a user or a third party and form opinions and valuable insight that saves time and mistakes for physicians, bypassing the bureaucracy and uncertainty of most strained medical industries.

The functions of AI serve not only to accelerate existing workflow patterns but to redefine the nature of clinical roles. In addition to treatment personalization, the growing emphasis on diagnostic precision allows professionals to devote more attention to addressing patient communication, ethical concerns, conclusion drawing, and nuanced decision-making that cannot yet be automated. While not a substitute for proper medical procedure, the multimodal data analysis conducted by machine learning can assist practitioners in predicting test outcomes and actionable results, minimizing the physical and emotional burden experienced by a patient during medical scans. For example, AI algorithms can analyze longitudinal patient data extracted from blood tests to detect early warning signs, monitor disease progression, and predict treatment risk (Nayak, 2025). This form of application in clinical support systems helps to mitigate the

growing reliance on laboratory testing, and offers an additional perspective to doctors. Furthermore, the analysis of laboratory test results is a difficult process that requires much expertise in the respective field. The necessity for certain specialists can greatly hinder diagnostic efficacy, particularly in rural or underdeveloped regions where tests may be conducted without the presence of the appropriate expert. An advanced ML program in future can help oversee and conclude the administering of such medical exams.

Limitations and Drawbacks

As seen throughout other sections of this review, AI-Driven algorithms in medical lab results offer numerous benefits; however, the pitfalls and consequences of their continued implementation must be equally acknowledged.

I.   Bias in Medical Algorithms

Often trained on unrepresentative datasets, AI models can exhibit biases, leading to disparities in healthcare outcomes (Obermeyer et al., 2019). This may be problematic when models are designed specifically to negatively target specific groups. (Obermeyer et al., 2019) Addressing these biases requires diverse and inclusive training datasets, strengthened validation processes, and constant monitoring of the AI systems. (Obermeyer et al., 2019)

II.   Data Privacy & Security Risks

Healthcare data is a prime target for cybercriminals, and the adoption of AI in healthcare has increased the volume and variety of data being generated, processed, and stored. (Marshall) Artificial intelligence models rely on vast amounts of patient data, which raises concerns about data privacy and security. (Marshall) Ensuring compliance with regulations such as HIPAA (Health Insurance Portability and Accountability Act) and the GDPR (General Data Protection Regulation) is essential to protect patient information. (Marshall) Secure data-sharing frameworks, as well as encryption methods must be implemented to prevent unauthorized access and breaches, and maintain patient confidentiality and trust. (Marshall) Ignorance of efficient data security could lead to data breaches, and hackers can easily manipulate data fed to AI for incorrect results (AI system vulnerabilities). Adversarial and ransomware Attacks continue to linger as threats. Privacy concerns involve data anonymization challenges and patient consent and ownership issues. (Marshall)

III.   Lack of Transparency

Based on a survey from Frontiers in Digital Health, transparency scores ranged from 6.4% to 60.9%, with a median of 29.1%. (Jonker at al.) They concluded that the major transparency gaps included missing documentation on training data, ethical considerations, and limitations for deployment. (Jonker at al.) The best way to close these gaps is by disclosure. It includes documenting and sharing the AI algorithm's logic and reasoning, the data inputs used to train the model, the methods used for model evaluation and validation and more. (Jonker at al.

IV. Over-Reliance on AI

Fears arise that doctors and hospitals will depend too much on AI tools and lose key human-based judgement. Modern predictive platforms like Pillar-0 and IDx-DR are highly accurate, with studies typically yielding an AUC value between 0.85-0.95 (Shuaib, 2020), but are not guarantees of diagnostic success nor surrogates for human thinking. At the current pace of AI development, it is not improbable that the point of technological singularity (TS), at which human intellect is completely surpassed by artificial capabilities, will lead to the diminishing or at least the augmentation of the human role in medicine (Shuaib, 2020). Though speculative, bias against AI implementation for this concern should remain a consideration.

V. Diagnostic Errors

An AI system has the potential, especially once highly integrated with a hospital or clinical environment, to make misclassifications and diagnostic errors. In coupling with the idea of human overreliance on artificial platforms, the likelihood of the occasional failure by an AI to correctly identify an affliction is all but guaranteed. A study published in the Journal of the American Medical Association found that ChatGPT-4, tested against 50 physicians, achieved a diagnostic accuracy of 90% against the doctors' average of 74% (Al-antari, 2024). A specialized model would likely perform even better than a general LLM like ChatGPT-4, but there remains a significant failure rate percentage that cannot be a substitute for human input. While effective as a secondary tool to support or contradict an assessment, AI is not yet at the stage of development nor public comfortability to be the sole arbiter of diagnostic decision making.

VI. Job Displacement

Concerns about the replacement of healthcare professionals by artificial intelligence (AI) have increased as AI capabilities expand. Although AI has shown effectiveness in areas such as diagnostics, data analysis, and robotic-assisted procedures, research indicates that it primarily functions as a supportive tool rather than a substitute for human clinicians. Studies demonstrate that AI systems can improve diagnostic accuracy and reduce administrative workload when used alongside physicians, while final clinical decision-making remains reliant on human judgment (Sharma, 2024). Additionally, organizations such as the World Health Organization emphasize that ethical reasoning, empathy, and contextual understanding cannot be replicated by AI (Sharma, 2024). As a result, the adoption of AI is expected to transform healthcare roles and create new positions focused on AI oversight and data management, rather than significantly reducing healthcare employment.

VII. High Costs of Implementation

The high cost of implementing artificial intelligence presents a significant barrier to its widespread adoption in healthcare systems. Developing and deploying AI technologies requires significant financial investment in advanced infrastructure, high-quality data collection, system integration, and ongoing maintenance. Research indicates that expenses related to training

algorithms, upgrading digital systems, and ensuring cybersecurity can place considerable strain on healthcare budgets, particularly for smaller hospitals and underfunded institutions (Sharma, 2024). Furthermore, additional costs associated with staff training, regulatory compliance, and ethical oversight increase the financial burden of AI adoption (Shuaib, 2020). Thus, high implementation costs may limit equitable access to AI-driven healthcare innovations.

VIII.   Regulatory & Legal Challenges

The meteoric rise of AI in the global healthcare sector has brought on regulatory and legal concerns, particularly in the question of liability. In Canada, for example, Health Canada classifies most clinical AI tools as 'software as a medical device' (SaMD), which practitioners may use at their own discretion but must still take responsibility for any decisions made (Kocabas, Bilgic, Gorgy, Harley, 2021). However, since the failure of the Artificial Intelligence and Data Act (AIDA) to pass through Parliament, no federal bill has since passed to overhaul largely outdated AI laws (Kocabas, Bilgic, Gorgy, Harley, 2021). Provincial statutes, federal councils, and quasi-independent government regulators are currently the predominant authorities on the issue. In the United States, AI systems are also regulated rather loosely, with state bylaws and federal agencies taking most authority. AI is not permitted to diagnose a patient independently, and the liability still rests on the healthcare provider for any faults carried out against a patient by an algorithmic mistake (Davtyan, 2025). As an official diagnosis must be conducted by an appropriately licensed medical professional, even the most advanced algorithm cannot substitute a doctor's judgement (Davtyan, 2025). Over-reliance on AI for clinical decision making, or a lack of informed consent by a patient that AI may influence their treatment, can result in lawsuits against the healthcare provider for negligence and medical malpractice (Davtyan, 2025). While these examples seem to draw relatively clear lines as to the liability of medical practitioners in their use of machine algorithms, they are under constant review and subject to change by governments as the rise of artificial efficiency becomes increasingly likely to surpass their human counterparts.

IX.   Patient Trust Issues

Patient trust remains one of the biggest challenges in the adoption of artificial intelligence within healthcare. Many patients express concerns about data privacy, algorithmic transparency, and the potential for errors in AI-assisted medical decisions. Studies indicate that limited understanding of how AI systems operate can reduce patient confidence, particularly when decision-making processes lack explainability (Witkowski et al., 2024). Additionally, documented cases of algorithmic bias in healthcare applications have raised concerns about fairness and equity, further undermining trust among patients (Witkowski et al., 2024). As a result, insufficient patient trust may hinder the effective integration of AI technologies into clinical practice.

Ethics

I. Patient Data Use and Privacy

AI diagnostic software hinges on extensive amounts of sensitive patient information, such as imaging scans, lab data, and electronic medical records (Price & Cohen, 2019). The magnitude and complexity of patient data make it even more vulnerable to data theft, unauthorized use, and improper distribution of personal medical information (Price & Cohen, 2019). With larger healthcare data systems, conventional privacy measures do not suffice in guarding against data aggregation on a massive scale and improper data reuse for tasks unrelated to patient treatment (Price & Cohen, 2019). The integration of AI must therefore be handled properly through proper data management, safe data storage, and proper data reuse constraints (Price & Cohen, 2019).

II. Informed Consent and Patient Autonomy

The incorporation of AI into diagnostic decision making affects patient autonomy by influencing medical outcomes that patients may not fully understand (McCradden et al., 2020). The use of artificial intelligence in diagnosis and treatment planning is often not transparent, as it operates in the background and beyond the patient's understanding and control (McCradden et al., 2020). In order to be medically ethical, patients should be made aware how AI systems could be part of their medical treatment (McCradden et al., 2020). It is unethical to unknowingly involve AI technology in a patient's care without transparency and proper explanation about how and when the technology will be used in treatment and affect their health (McCradden et al., 2020).

III. Algorithmic Bias and Fairness

AI diagnostic systems are trained on historical medical data that may not represent the full diversity of patient populations (Obermeyer et al., 2019). When there is imbalance in training data, there might be lower accuracy in the diagnosis provided by the AI system to particular demographics (Obermeyer et al., 2019). Due to disparities, there might be discrepancies in the diagnosis, resulting in increasing the inequities in the diagnosis process of healthcare services (Obermeyer et al., 2019). Ethical usage of AI demands constant performance evaluation of these tools with respect to various demographics (Obermeyer et al., 2019). Corrective actions have to be taken once bias is found, which can be detrimental to particular demographics (Obermeyer et al., 2019).

IV. Human Oversight and Accountability

Artificial intelligence systems used in diagnosis can sometimes lead to errors and should never be used as autonomous decision makers. The overdependence on suggestions could create complacency among doctors and may lead to errors (London, 2019). The responsibility to make decisions should always lie with licensed medical professionals, irrespective of the levels of accuracy an AI may possess (London, 2019). With human oversight, decisions related to patients are guaranteed to continue emphasizing contextual thought and ethical responsibility (London,

2019). The integration of ethical considerations with AI relies on maintaining the secondary nature of AI when it comes to medical decisions (London, 2019).

Future Directions

With all considerations, it is evident that determining the future of AI's role in healthcare is the next step in its evolution. While proven to be extremely adept in tasks suited to machine learning and natural processing, artificial systems also bring about uncertainties relating to professional displacement, legal and financial fragility, and ethical questions that remain unsolved. Nonetheless, the pace of AI development, both in its integration into healthcare and other fields and in its exponentially growing investment worldwide, does not seem likely to slow down. Rather, it is more plausible that the increasing accuracy of ML modeling will eventually make it indispensable to professionals in diagnostic and data analytics, as the efficacy of AI has already and will continue to further surpass any form of human ability in this capacity.

However, contrary to fears of a shrinking job market in healthcare due to such artificial encroachment, the role of medical practitioners will shift to accommodate for tasks AI cannot do (Sharma, 2024). While the need for dedicated radiologists and pathologists may decline, the demand for experts capable of interpreting, managing, and directing AI systems will rise in a negative correlation (Sharma, 2024). It is more predictable that, like in most industries, the introduction of a new technology will alleviate the burden of menial tasks and allow for redirection to uniquely human functions, such as emotional communication, ethical and moral decision-making, or creative understanding (Sharma, 2024). Therefore, it is necessary for medical training programs and education to adapt to evolving conditions within the field, and offer aspiring future doctors to become familiar with AI management and a more interpretive form of critical analysis. Rather than shun the inevitability of AI intertwining with clinical and pharmaceutical settings, the focus should shift towards ensuring the tool is used ethically and responsibly, primarily through an emphasis on professional education on the subject and overarching government/corporate regulations to prevent private overreach and liability concerns.

Conclusion

With the integration of artificial intelligence in algorithmic data, analytical software, and pharmaceutical development, the modern clinical industry has made significant strides towards the goal of wholly accurate and interpretable medical prescription and diagnosis. Current AI models, which may suit tasks ranging from the specific identification of an affliction to general imagery analytics or drug trial hypothesis, have already been tested to be on par and often more efficient in making accurate, unbiased decisions compared to human experts (Shuaib, 2020). Where these systems show to be particularly useful is in the rapid organization and decipheration of enormous complex data sets, with popular platforms like Aidoc becoming commonplace in hospital settings worldwide for their utility in data interpretation. The pace and scale at which machine learning is able to consolidate heterogeneous data not only allows for more efficient

conclusions to be drawn, but also increases the likelihood of further interconnected insights and problems to be solved, exemplified by the breakthroughs in novel drug development made by Insilico Medicine's predictive system (Yi, 2018). While boasting undeniable clinical benefits, artificial intelligence is equally the target of much criticism and concern, especially in a field as crucial and subject to evolution as medicine. Transparency, privacy, and regulatory issues relating to the hasty implementation of AI systems in recent years may continue to sow doubt on the reliability and ethics of these new platforms, most of which are privately-owned and have independent data storage. Worry about potential bias present in AI training and the possibility of diagnostic errors also leads to questions on liability and public distrust. These concerns are primarily related to novelty and lack of experience with these foreign entities, so ensuring that a transparent standard of ethical practice is maintained and enforced by governments and independent regulators (ex. FDA, Health Canada) amidst the integration of machine learning into clinical diagnosis is important in maintaining public support.

Works Cited

Akhloufi , M. A., & Jandoubi, B. (2025). Multimodal Artificial Intelligence in Medical
    Diagnostics. MDPI; Information.
    https://drive.google.com/file/d/1r62S4KcuqIatjArPUikppb4apevj0iU3/view

Al-antari, M. A. (2024). Integration of AI and data-driven algorithms in diagnostic software.pdf.
    MDPI; Diagnostics.
    https://drive.google.com/file/d/1CSQyU6tKJsDpIfF1H2heB0p66tXUCgYc/view

Bharadiya, J. P. (2023). Rise of Artificial Intelligence in Business and Industry. Journal of
    Engineering Research and Reports, 25(3), 85–103.
    https://d1wqtxts1xzle7.cloudfront.net/103844778/1776-libre.pdf?1687972946=&respons
    e-content-disposition=inline%3B+filename%3DRise_of_Artificial_Intelligence_in_Busi
    n.pdf&Expires=1767151821&Signature=Ym9xAuYbpPJ~zx4T-3rKmbmwih6yj1gIw7pJ
    mk46j6TG3ZolydkPE4e43PDEHj7~44MJrp~XzRXDnqJiQRPtc2sg4ZmI8o-d5hlvtyJBvf
    HdLwWBq1wBVPA1MGozgv9HhQNa4Dyvsqtx98MrM4QZCD9iF9P~-lx10A2Y5~dy
    UjPFQoyNvKjyqH9YeMsknSp2cciwBGYoUXrRrI1kNWRGDwaWea8gs74vMhpNbDI
    Xp9bdmV2JYHth54sYRGg0YbuUENhTNm23ADwJtefiufrbDeqQSc6ScPcKbTArLwya
    Luk~1167Q~a~vQ50BOgZj3zRHkzedTZSMwtlarG8YB19LQ__&Key-Pair-Id=APKAJ
    LOHF5GGSLRBV4ZA

El Arab, Rabie Adel, and Omayma Abdulaziz Al Moosa. "Systematic review of cost
    effectiveness and budget impact of artificial intelligence in Healthcare." Npj | Digital
    Medicine, vol. 8, no. 1, 26 Aug. 2025, https://doi.org/10.1038/s41746-025-01722-y.

Hazarika, Indrajit. "Artificial Intelligence: Opportunities and implications for the Health
    Workforce." International Health, vol. 12, no. 4, 17 Apr. 2020, pp. 241–245,
    https://doi.org/10.1093/inthealth/ihaa007.

Hirosawa, Takanobu, and Taro Shimizu. "A narrative review of Artificial Intelligence in medical
    diagnostics." Computers, Materials & Continua, vol. 83, no. 3, 2025, pp.
    3919–3944, https://doi.org/10.32604/cmc.2025.063803.

Journal of Law, Technology, & the Internet. (2025). The U.S. Approach to AI Regulation:
    Federal Laws, Policies, and Strategies Explained. Journal of Law, Technology, & the
    Internet, 16(2).
    https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?article=1172&context=jolti

Khan, Z. et al. (2025). Diagnostic Accuracy of IDX-DR for Detecting Diabetic Retinopathy: a
    Systematic Review and Meta-Analysis. American Journal of Ophthalmology, 273(May
    2025), 192–204. https://doi.org/10.1016/j.ajo.2025.02.022

Kocabas, S. et al. (2025). Deconstructing Canada's efforts to integrate artificial intelligence in
    medicine and medical education. McGill Journal of Medicine, 19(1).
    https://doi.org/10.26443/mjm.v19i1.871

London, A. J. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus
    Explainability. Hastings Center Report, 49(1), 15–21. https://doi.org/10.1002/hast.973

Marshall, S. (2024, September 24). AI in Healthcare: Security and Privacy Concerns. Lepide Blog: A Guide to IT Security, Compliance and IT Operations. https://www.lepide.com/blog/ai-in-healthcare-security-and-privacy-concerns/

McCradden, M. D., Joshi, S., Anderson, J. A., Mazwi, M., Goldenberg, A., & Zlotnik Shaul, R. (2020). Patient safety and quality improvement: Ethical principles for a regulatory approach to bias in healthcare machine learning. Journal of the American Medical Informatics Association, 27(12), 2024–2027. https://doi.org/10.1093/jamia/ocaa085

Nayak, B. S. (2025). The Evolution and Architecture of Multimodal AI Systems. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11(1), 1007–1017. https://doi.org/10.32628/cseit251112108

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Price, W. N., & Cohen, I. G. (2019). Privacy in the Age of Medical Big Data. Nature Medicine, 25(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7

Russell, S., & Norvig, P. (2022, August 22). Artificial Intelligence: A Modern Approach, 4th US ed. Aima.cs.berkeley.edu. https://aima.cs.berkeley.edu/

Sharma, M. (2024). The Impact of AI on Healthcare Jobs: Will Automation Replace Doctors. American Journal of Data Mining and Knowledge Discovery, 9(2), 32–35. https://doi.org/10.11648/j.ajdmkd.20240902.11

Shuaib, A. et al. (2020). The Increasing Role of Artificial Intelligence in Health Care: Will Robots Replace Doctors in the Future? International Journal of General Medicine, 13(2020), 891–896. https://doi.org/10.2147/ijgm.s268093

Witkowski, K et al. (2024). Public perceptions of artificial intelligence in healthcare: ethical concerns and opportunities for patient-centered care. BMC Medical Ethics, 25(1). https://doi.org/10.1186/s12910-024-01066-4

Yi, F. et al. (2018). In silico approach in reveal traditional medicine plants pharmacological material basis. Chinese Medicine, 13(1). https://doi.org/10.1186/s13020-018-0190-0

Zaazoue, K. A. et al. (2023). Evaluating the Performance of a Commercially Available Artificial Intelligence Algorithm for Automated Detection of Pulmonary Embolism on Contrast-Enhanced Computed Tomography and Computed Tomography Pulmonary Angiography in Patients With Coronavirus Disease 2019. Mayo Clinic Proceedings: Innovations, Quality & Outcomes, 7(3), 143–152. https://doi.org/10.1016/j.mayocpiqo.2023.03.001

Ethics of Artificial Intelligence in Mental Healthcare By Courtney Chow

Abstract

Artificial intelligence is becoming increasingly integrated into mental healthcare, raising complicated ethical questions about trust, empathy, privacy, and clinical responsibility. This paper examines both the potential and the limitations of AI in mental healthcare. Critics argue that AI cannot genuinely empathize, risks over-reliance among vulnerable populations, and introduces significant privacy and safety concerns in crisis situations where human judgment is essential. Supporters highlight how digital therapeutics and automation may expand access, reduce costs, and ease the strain on overburdened mental health systems. Generational and cultural differences further complicate public attitudes toward AI, with younger digital natives showing greater acceptance than older individuals who typically prioritize spiritual, interpersonal, and traditional forms of support. Through an analysis of current research, clinical concerns, and societal perspectives, this paper argues that while AI may serve as a helpful supplement, it lacks the emotional, moral, and relational capacity required for primary therapeutic care. Ensuring safe and ethical integration of AI in mental health requires strong oversight and serious commitment to preserving the human connection at the center of healing.

Introduction

Artificial intelligence has crept into nearly every corner of modern life, from helping students with essays to answering questions faster than a Google search. Unsurprisingly, psychology is now part of the experiment. The debate is fierce: can a machine genuinely empathize, or is it just a well-programmed actor reciting lines? Critics argue that AI raises privacy concerns, encourages over-reliance on technology, and cannot truly understand human emotion. Supporters, meanwhile, highlight how digital therapeutics and automation can improve efficiency and provide much-needed support to overburdened professionals.

Literature Review

Critics remind us that therapy is not the same as ordering groceries online. Mental health treatment is built on trust, emotional understanding, and empathy, qualities that are impossible for AI to authentically replicate. While AI can analyze tone, facial expressions, or even the length of a pause in conversation, opposers argue that it is simply imitating understanding rather than genuinely feeling." AI does not partake in emotional experiences-it neither shares joy nor sorrow. Therefore, regardless of how eloquently it crafts a response to seem like it shares an emotional experience, this response will be untruthful, as it does not share any experience."(Rubin). A patient may pour out their fears and grief, but unlike a trained professional, a computer will never truly "sit with them" in that pain. This lack of authentic emotional connection raises concerns that patients may not receive the depth of care they need.

Another concern is over-reliance on technology. People already depend heavily on AI for school, work, and daily tasks. Extending and normalizing that dependence into mental health

care could be dangerous. Unlike homework, where a wrong answer means a bad grade, mental health mistakes can have lasting effects. Therapy requires accountability, adaptability, and genuine human connection, all of which are human strengths that machines cannot replicate. Healing is not simply an exchange of information; it is the feeling of being understood, validated, and seen.

Still, some of the strongest arguments in favor of AI in mental health revolve around digital therapeutics and the efficiency they bring. These tools have shown measurable success in treating depression, anxiety, PTSD, ADHD, and substance use disorders. Many combine cognitive-behavioral therapy principles with interactive technology, allowing patients to access evidence-based treatment whenever they need it. Their credibility isn't just hype: the FDA has approved two mental health digital therapeutics, signaling that these tools are clinically validated and more than experimental gadgets (Rostami). Critics respond that, regardless of technological success, emotional nuance, intuition, and empathy remain central to therapy and cannot be replicated.

Then, there is the glaring issue of privacy. Doctors and psychologists are bound by strict oaths and legal protections such as HIPAA or basic ethical principles of confidentiality, designed to keep sensitive information confidential. AI, however, doesn't swear loyalty to its patients, it runs on servers, algorithms, and updates. While developers are making strides in encryption and compliance, critics remain skeptical. After all, "trust me, we encrypted it" doesn't carry quite the same comfort as a doctor saying, "I am legally bound never to repeat this." For many, the thought of confiding personal trauma to a machine sparks more paranoia than relief. According to the article Security Implications of AI Chatbots in Health Care, "Data can be disclosed to any intended and unintended audiences and used for various purposes without authorization" (Li). That being said, patients who already struggle to trust people won't suddenly trust a glowing screen.

Discussion

1. Generational Differences in Trust and Acceptance

While much of the older generation displays scepticism on the subject of AI, younger people may approach the technology differently. Perhaps the first generation in history which can truly be considered "digital natives"; having grown up with smartphones, social media, and virtual assistants, they are generally more comfortable interacting with AI and may view it as a convenient, familiar tool, rather than an intimidating replacement for human care. Cost also plays an important role in this divide. Many older adults have more financial stability or savings that allow them to afford in-person therapy, while younger generations often face financial constraints that make traditional therapy less accessible. As a result, they may turn to AI-based tools because they are more affordable and available around the clock. However, this economic convenience can lead to an overreliance on technology for emotional support, a serious and growing concern in a world already shaped by digital dependency. Ignorance of or

misunderstanding such limitations could lead to a therapeutic misconception (TM) where an individual would underestimate the restrictions of such technologies and overestimate their ability to provide therapeutic support and guidance."(Kawaja). Relying too heavily on AI for mental health care may blur the line between genuine human connection and automated responses, potentially weakening emotional resilience and interpersonal skills over time. This reinforces the idea that AI should serve as a supplement to mental health care, not a replacement for the authentic human support that is essential to healing.

2. Cultural and Religious Factors

Religion adds another layer to this generational divide. For many older adults, churches, mosques, temples, and other places of worship have traditionally been where people turned for counsel, reflection, and community support. In contrast, AI offers a type of guidance stripped of spiritual meaning or moral depth. For those deeply rooted in spiritual traditions, the rise of AI could feel like a challenge not only to human connection but to divine connection as well. According to Jeff Cickell of the University of Chicago, studies have shown that simply reading about AI had a "stronger negative effect on participants' religious convictions than reading about other kinds of breakthroughs, even when the breakthroughs were judged as equally impressive". This fundamental difference makes it difficult for religious individuals to trust AI in areas concerning mental or emotional well-being. Moreover, the rise of digital therapy and guidance invites an unsettling question: if people begin to seek emotional comfort from technology rather than faith, what role remains for religion in a world increasingly shaped by machines? After all, there is a significant difference between pouring your feelings out to your pastor and talking to an inanimate machine that you know has no capability to truly care about your problems!

3. Crisis Responsibility and Accountability

The increasing use of AI in mental health care also raises troubling questions about responsibility in moments of crisis. What happens if someone experiencing suicidal thoughts or self-harm relies on an AI program for help and the system fails to respond appropriately? In traditional therapy, clinicians are trained to detect subtle warning signs, assess risk, and take immediate steps to ensure a patient's safety. They can make judgment calls grounded in empathy, intuition, and moral responsibility while AI cannot. It processes data rather than emotion, pattern recognition rather than moral urgency. "However, if a (mental health) app provides false information or responds inadequately, it may not only create a false sense of security but actually deepen the crisis." (Engelson) If an individual in distress turns to a chatbot or digital therapeutic tool and receives an unhelpful, delayed, or even harmful response, the consequences could be catastrophic. Yet, the question of who bears responsibility remains unresolved. Should blame fall on the developers who designed the algorithm, the institutions that deployed it, or the users who placed their trust in it? At present, there are few legal or ethical frameworks to guide these situations. Developers often position their products as "tools" rather than substitutes for professional care, while healthcare systems may lack the capacity to supervise every digital

interaction). This creates a dangerous accountability gap; one in which a vulnerable person could be failed by technology without anyone legally or morally answerable for the harm. The stakes are particularly high in mental health, where a single missed cue or impersonal response can quite literally make the difference between life and death. Without crisis protocols, such as real-time human monitoring, escalation pathways, and external oversight, AI-assisted therapy risks prioritizing efficiency over safety. Innovation in mental health care should never come at the expense of compassion, protection, or the sanctity of human life.

Conclusion

While I acknowledge that critics raise valid points about privacy, emotional authenticity, and over-reliance, these concerns outweigh the optimism surrounding AI's potential. Artificial intelligence may enhance accessibility and efficiency, but mental health care is not simply a system to be optimized. It is a relationship built on trust, compassion, and human understanding. AI can help identify patterns or provide resources, but it remains fundamentally limited in its ability to comprehend the depth of human suffering. Though AI may one day become a powerful aid, it is not ready to fully take over such a personal and ethical domain. The goal of AI in mental healthcare should not be for AI to replace or redefine therapy, but to support and strengthen the uniquely human connection that makes healing possible.

Works Cited

Cickell, Jeff. (2024). "Where AI Thrives, Religion May Struggle." Chicago Booth Review, 26 Mar. 2024, https://www.chicagobooth.edu/review/where-ai-thrives-religion-may-struggle.

Engelson, Shlomo. (2025). "Dangerous AI Therapy-Bots Are Running Amok. Congress Must Act." The Hill, https://thehill.com/opinion/technology/5373802-ai-therapist-bots-pose-risks/#:~:text=The %20regulatory%20questions%20are

Khawaja, Zoha, and Jean-Christophe Bélisle-Pipon. (2023). "Your Robot Therapist Is Not Your Therapist: Understanding the Role of AI-Powered Mental Health Chatbots." Frontiers in Digital Health, article 1278186, doi:10.3389/fdgth.2023.1278186. https://journalfph.com/index.php/jfph/article/view/9

Li, Jingquan. (2023). "Security Implications of AI Chatbots in Health Care." Journal of Medical Internet Research, vol. 25, 28, e47551. JMIR Publications, doi:10.2196/47551. PMCID: PMC10716748. https://www.jmir.org/2023/1/e47551/

Rostami, Mehdi, Mansour Abdi, and Nazanin Zahra Asadi. (2025). "Predicting the Future of Mental Health Interventions: The Role of Digital Therapeutics." Journal of Foresight and Health Governance, vol. 2, no. 2, pp. 33-51. https://journalfph.com/index.php/jfph/article/view/9

Rubin, M., Arnon, H., Huppert, J., & Perry, A. (2024)."Considering the Role of Human Empathy in AI-Driven Therapy." JMIR Mental Health, vol. 11, article e56529, doi:10.2196/56529. https://pubmed.ncbi.nlm.nih.gov/38861302

How Short-Term Moderate Aerobic Exercise Affects Adolescent Cognitive Function
By Brandon Wong

Abstract

Short bouts of exercise can cause temporary changes in the brain that improve how we think and function. Past research in this field shows that reaction time does not consistently improve with acute exercise, and in some cases may even decline during physical activity. In contrast, executive function -like self-control, inhibition, and flexible thinking shows the most reliable and significant improvements after 20-35 minutes of moderate-intensity exercise. Findings for memory fall between these two patterns: short-term and working memory show some benefits, though the results are less consistent than those observed for executive function. Most of these studies, however, have been done with adults, not teenagers. Newer studies suggest that these mental boosts may come from things like better blood flow to the brain and the release of brain chemicals such as brain-derived neurotrophic factor (BDNF), which supports learning and memory. These chemicals create a temporary "primed" state that enhances information processing and memory. This study tests whether a single 20–30 minute short session of moderate aerobic exercise helps healthy adolescents perform better on tasks measuring reaction time, executive function, and memory compared to a group that rests quietly. Executive function includes self-control and flexible thinking, which are critical for decision-making in school settings.

Introduction

A growing body of research suggests that aerobic exercise can temporarily boost brain function. Studies in children and adults have shown improvements in reaction time, learning, and working memory following short bouts of moderate-intensity activity lasting 20–35 minutes. (Garrett et al., 2024) Reaction time, executive function, and memory are three cognitive domains closely tied to both academic performance and everyday problem-solving. The improvements are thought to stem from increased cerebral blood flow, improved oxygen delivery, and the release of neurotransmitters and neurotrophic factors like BDNF, which support learning and memory formation. However, far fewer studies have examined whether adolescents experience similar benefits. Because this age group is still undergoing brain maturation, particularly in the prefrontal cortex, it is unclear whether the effects of acute exercise will resemble those found in adults or show unique patterns. Understanding these effects has practical importance: If a single session of exercise improves cognition in adolescents, it could support the integration of short exercise breaks into school schedules to enhance learning, attention, and academic performance.

Literature Review

Whereas other studies report cognitive gains with exercise, the general trend of studies reports little to no increase in reaction time following exercise. For instance, Brisswalter has reported that the performance on tests of cognition can actually be decreased during acute

exercise intervals, but better preservation occurred with higher levels of fitness (Brisswalter et al. 1997). Reddy, Eckner, and Kutcher also examined the clinical measure of reaction time, RTclin and reported that performance improved through repeated practice but not through exercise per se (Reddy et al. 2014). Similarly, Chandra's. report that although exercise by itself did not cause significant changes in reaction time, environmental heat stress reduced attention and speed of response during exercise (Chandra et al. 2010). Such results indicate that acute exercise may not always improve reaction time, and, on occasion, performance on the task can be reduced through the competition between the physical and the task demands for attention.

Evidence on the effect of acute exercise on the performance of executive function demonstrates a consistent association between exercise and enhanced inhibitory control, task shifting, and decision-making. Research conducted on children and adolescents also reveals the maturation of executive function. For instance, the results by Yuan-Shuo Chan et al. (2021) uncovered significant improvements to the executive functioning of children with ADHD after acute bouts of exercise and thus suggested that exercise is an effective behavioural intervention (Chandra et al. 2010). Likewise, an earlier systematic review exposed that aerobic exercise stimulated the greatest effects on the performance of executive function among children with ADHD, although other exercise methods also stimulated significant increments (Zhu et al. 2023). Extending on this, a comprehensive review by Zhu et al. (2024) calculated that the participants with ADHD showed the largest enhancements to the performance of the executive function among children and adolescents, who showed more increments compared to adults. Most importantly, however, these increments turned out to be corresponding to interventions that comprised moderate to low and not high intensity exercise (Singh et al. 2025). By themselves, these outcomes indicate that exercise—particularly aerobic and moderate intensity exercise— enhances the performance of the executive function, especially among youths with ADHD.

Physical activity has also been associated with enhanced memory, although the evidence is inconsistent by age group and by the kind of memory tested. Acute exercise interventions improved working memory and recall among children with attention-deficit/hyperactivity disorder and thus hold the potential to be an adjunctive cognition support mechanism among this group (Yang et al., 2021). In another study, Singh (2025) examined the data on acute and chronic exercise. They concluded that exercise has a strong beneficial effect on memory among young to middle-aged adults, but that the evidence is more restricted among adolescents (Singh et al. 2025). Laboratory studies among children also provided evidence of an immediate benefit: moderate aerobically intensive exercise improved verbal learning and short-term recall, but its effect on long-term but not short-term recognition was uncertain (Etnier, 2013).

Discussion

This analysis examines how acute aerobic exercise influences cognitive performance— specifically reaction time, executive function, and memory— and finds that while exercise consistently benefits executive control, its effects on reaction time and memory are less reliable. Across research studies, aerobic and moderate exercise intensity has reliably improved executive

function and limited memory. However, results for reaction time are inconsistent and sometimes either do not improve or temporarily worsen performance during exercise. Such inconsistencies can be due to differences between task type, exercise intensity, and participants' fitness levels.

One of the overriding trends to come out of the studies is the potent effect observed at the younger ages, particularly children and adolescents, on working memory and executive function. Stroop effect task research on the exercise-induced gains in individuals with diagnosed ADHD documents particularly strong gains and suggests that exercise can facilitate improved inhibitory control and task-switching performance through heightened arousal and neurochemical activity. By comparison, the adult and older adult studies that dominate the field have examined long-duration exercise interventions yielding benefits on memory and cognitive maintenance but not on short-duration, instantaneous gains.

These results indicate that the inclusion of ordinary exercise at an early age could produce not only short-term but also long-term benefits on cognition. For adolescents, short durations of moderate-intensity aerobic exercise can potentially boost concentration and academic-executive function. Also, the maintenance of an active life can potentially aid in the delay or prevention of age-related cognitive decline. This highlights the potential future preventive value in exercising at an early age—both for physical health and also for the long-term maintenance of cognitive function. Despite these promising results, an important gap remains between the available research directly aimed at adolescents. Most evidence available is drawn from adult populations and thus places limitations on our understanding of how developmental considerations relate to the cognition-linked benefits of exercise.

Taken together, these studies indicate that short-term recall and memory are highly sensitive to acute exercise, but long-term consequences may be moderated by age and by repeated exposure to exercise.

Relationships and Implications

The cumulative findings from various studies indicate that exercise positively influences cognition through several pathways, namely biological (enhanced cerebral blood flow and BDNF release), psychological (increased motivation and alertness), and behavioural (better task engagement). Incorporating brief exercise sessions into school timetables or study practices may lead to significant improvements in attention and learning outcomes. Furthermore, recognising the role of early-life exercise in promoting lifelong brain health could guide public health initiatives designed to prevent cognitive decline among older populations.

Limitations

While this analysis highlights consistent patterns between exercise and cognitive performance, several limitations should be acknowledged. Although this review uses findings from existing studies rather than original experimental data, it can be used to identify consistent patterns across research. Variations in study design, demographics, and cognitive testing methods may influence outcomes, but they also highlight possibilities for future studies.

Many of the studies reviewed focused on short-term or acute effects, making it difficult to determine whether improvements in executive function or memory persist over time. Research that tracks participants across developmental stages would help explain whether exercise drives lasting cognitive growth or temporarily enhances performance.

Future studies should consider the dynamics by which these benefits take effect in younger groups and the correlations between exercise intensity, duration, and the process of successive brain maturation. Experimental studies measuring different age groups could clarify whether exercise has an influential effect on the process of developing cognition or has an effect mainly on existing capabilities.

Currently, most research still emphasises adult or clinical populations, leaving a gap in understanding how exercise impacts typically developing adolescents. More work is needed to explore how variables like exercise intensity, duration, and frequency affect the process of brain maturation. Focusing on these limitations will provide a clearer picture of how exercise influences cognitive development and how it can be most effectively used to improve youth health and education programs.

Works Cited

Bathina, Siresha, and Undurti N. Das. "Brain-derived neurotrophic factor and its clinical implications." PubMed Central, 11 Dec. 2015, https://doi.org/10.5114/aoms.2015.56342. Accessed 26 Sept. 2025.

Brisswalter, J., et al. "Influence of physical exercise on simple reaction time: effect of physical fitness." PubMed, Dec. 1997, https://doi.org/10.2466/pms.1997.85.3.1019. Accessed 29 Sept. 2025.

Chan, Yuan Shuo, et al. "Effects of physical exercise on children with attention deficit hyperactivity disorder." PubMed, Apr. 2022, https://doi.org/10.1016/j.bj.2021.11.011. Accessed 28 Dec. 2025.

Chandra, Ananga Mohan, et al. "Effect of exercise and heat-load on simple reaction time of university students." PubMed, 2010, https://doi.org/10.1080/10803548.2010.11076862. Accessed 28 Dec. 2025.

Garrett, Jordan, et al. "A systematic review and Bayesian meta-analysis provide evidence for an effect of acute physical activity on cognition in young adults." Nature, 28 Aug. 2024, https://doi.org/10.1038/s44271-024-00124-2. Accessed 28 Dec. 2025.

Li, Ruei-Hong, et al. "Acute Concurrent Exercise Improves Inhibitory Control and Its Non-Mediation Role of Lactate: An Event-Related Potential Study." Research Square, 25 Jan. 2025, https://doi.org/10.21203/rs.3.rs-4278500/v1. Accessed 28 Dec. 2025.

Moriarty, Terence A., et al. "Acute Aerobic Exercise Based Cognitive and Motor Priming: Practical Applications and Mechanisms." Frontiers, 11 Dec. 2019, https://doi.org/10.3389/fpsyg.2019.02790. Accessed 25 Sept. 2025.

Reddy, Shailesh, et al. "Effect of acute exercise on clinically measured reaction time in collegiate athletes." PubMed, Mar. 2024, https://doi.org/10.1249/mss.0000000000000140. Accessed 29 Sept. 2025.

Singh, Ben, et al. "Effectiveness of exercise for improving cognition, memory and executive function: a systematic umbrella review and meta-meta-analysis." PubMed, June 2025, https://doi.org/10.1136/bjsports-2024-108589. Accessed 28 Dec. 2025.

Yang, Zongyu, et al. "The relationship between acute aerobic exercise and inhibitory control in college students: The impact of physical and cognitive engagement." ScienceDirect, 1 Mar. 2025, https://doi.org/10.1016/j.physbeh.2024.114779. Accessed 28 Sept. 2025.

Zhu, Feilong, et al. "Comparative effectiveness of various physical exercise interventions on executive functions and related symptoms in children and adolescents with attention deficit hyperactivity disorder: A systematic review and network meta-analysis." Comparative effectiveness of various physical exercise interventions on executive functions and related symptoms in children and adolescents with attention deficit hyperactivity disorder: A systematic review and network meta-analysis, Mar. 2023, https://doi.org/10.3389/fpubh.2023.1133727. Accessed 29 Sept. 2025.

Filling in the Lower Black Hole Mass Gap By Eoin Welsh

Abstract

This paper discusses black holes, discovered through gravitational waves, that fall within the previously theorized lower mass gap. It analyzes common trends between the seven objects, and theorizes about possible reasons objects in this range were not previously detectable. These objects had a mean chi efficiency near zero, suggesting they were all formed through star clusters, rather than originating as binary star systems. This explains why events such as these were previously undetectable, while also leaving new questions about their absence in other binary system formation methods.

Introduction

Black holes were first theorized to exist in the 16th century, but not detected until 1972. Initially, scientists saw the effects of them using X-ray binaries, binary systems containing one star and one black hole, resulting in electromagnetic waves that could be detected from Earth. These x-ray binaries detected black holes ranging from 5 $M_\odot$ to 23 billion $M_\odot$ over 43 years. Over time, our means of detecting them have determined what we can know about them.

To date, scientists thought there was a fundamental, universal property that meant stellar black holes couldn't form with less than five stellar masses. Before 2015, binary black hole and neutron star systems were primarily detected as X-ray binary systems, using electromagnetic waves [7]. In 2015, the first gravitational wave was detected by the Laser Interferometer Gravitational-Wave Detector (LIGO), and a new way of discovering these systems was available. To date, 218 confirmed gravitational waves have been detected by these systems, leaving scientists with a bigger data pool than ever, including many systems previously undetectable by previous methods [1]. However, this opens up a new opportunity to question these previous theories.

Most scientists agreed on the presence of a lower black hole mass gap, but until recently there was no other detection method readily available to confirm that this was the case. It could have very well been that there was instead a flaw in the x-ray binary detection method itself, something that couldn't be proven in either direction without gravitational waves. Do binary black hole and neutron star systems detected by gravitational waves still support the existence of a lower black hole mass gap? The lower black hole mass gap is a theory which proposed there was a fundamental factor of the universe that prevented black holes from forming below 5 $M_\odot$ which was greater than both the largest observed neutron star at 2.5 $M_\odot$ [8], and the highest theoretical neutron star mass at 3 $M_\odot$ [11]. This study aims to analyze the gravitational waves with theorized masses near or within this proposed gap- from 2.5 to 5 $M_\odot$, and determine if there is indeed a lower black hole mass gap, and if not, why these black holes weren't detected by x-ray binaries. This study aims to demonstrate if stellar mass black holes below 5 $M_\odot$ exist, and

if they do, why they remained undetected for so long, increasing our understanding of the range of black holes and their mass. First, I will review the relevant preexisting literature and address some of the gaps in our current knowledge. Then I will explain the new way of detecting black holes, and how it affects the theory of the lower mass gap. Then I will discuss findings and implications.

Literature Review

The lower black hole mass gap was originally theorized to exist due to the lack of detection of black holes below 4.3 $M_\odot$ through x-ray binaries [6]. Before gravitational waves were first officially detected, x-ray binaries were the primary way to detect black holes. The newer ability to detect gravitational ways offered scientists the opportunity to either strengthen or oppose the lower black hole mass gap, as gravitational waves didn't rely on electromagnetic radiation and therefore weren't subject to the same experimental limitations as electromagnetic-based detection methods. However, scientists have not yet tested this hypothesis using gravitational wave data, despite new advances in technology. In addition, if there is indeed a mass gap, then one must question why it is there. If there isn't a lower mass gap, then there is at least a dip in the distribution of stellar masses between 3 and 7, with a variety of potential explanations. One of the most plausible reasons is that black hole binaries between 3 and 7 stellar masses form in a different way to larger ones, a formation method that is rarer in the observable universe [10].

Neutron stars are the other piece of the puzzle in terms of merger systems. While less often detected through gravitational waves, due to their comparatively smaller size, they give off electromagnetic waves and therefore have been detectable for much longer. The alternative hypothesis could be that neutron stars fill this mass gap. There is no set upper limit for the mass of a neutron star, but studies agree that a rough limit can be placed at $3M_\odot$ [6, 11, 15]. Other studies observing the distribution of x-ray binaries have determined a mass gap is likely, and Farr and Straven give an approximate value for this gap of between 3 $M_\odot$ and 4.3 $M_\odot$ [9] To estimate this number, the data was split into two samples, one with only low mass systems, and one with both low and high mass systems. They ran a variety of different Bayesian statistical models on each, and used the model most accurate to each sample for analysis. The best models for the low mass and combined samples respectively evaluated the minimum black hole mass, calculated as the 1% quantile, at 4.3 $M_\odot$ and 4.5 $M_\odot$ respectively, both with 90% confidence. The paper proposes two possible explanations for the perceived gap. Either there is indeed some fundamental property of the universe preventing stellar masses from forming in that range, or there was a selection bias in the study preventing stellar masses within the 'gap' from being detected. At the time, researchers could only speculate, but with the advent of gravitational wave detection, free from some of the limitations of earlier methods, more definitive answers may now be within reach.

Gravitational waves as a concept were first proposed by Einstein in 1916, as part of his theory of general relativity. In the early 1960s interferometric detectors were first proposed, and soon after the turn of the century the first of such machines were constructed in various locations throughout the globe. Two of the most prominent of these are the Laser Interferometer Gravitational-Wave Observatories (LIGO) in Livingston, Louisiana, and Hanford, Washington. The machines were tested and improved throughout the decade, and on September 14, 2015, both LIGO observatories detected their first ever gravitational wave [1]. The henceforth dubbed GW150914 was verified on account of its strength compared with potential noise, the matching waveforms at both laboratories, and its matching of a predicted form for a gravitational waves signal. After analyzing the waveform and its corresponding data, researchers determined that GW150914 was from a binary black hole merger, and the subsequent ringdown of their combined remnant. When two black holes merge, they first have an inspiral where they steadily orbit closer together. Next is the merger, when the two black holes finally combine into one. This is when gravitational waves are primarily emitted. After the merge, the remaining combined black hole continues to wobble in place before settling into its new position, still orbiting the center of mass very slightly. This is called the ringdown. GWW150914 was the first concrete evidence for the existence of binary stellar-mass black holes, and the first time scientists had access to a way of observing the universe without relying on electromagnetic waves.

The paper by LIGO is the third comprehensive analysis of data from the first three LIGO observing runs, providing insight into the population of black holes across the universe [2]. As the number of confirmed gravitational wave detection's has finally reached a substantial enough number, additional predictions can now be made about the commonality of different black hole formation paths, and the average distribution of many measured qualities. A Bayesian model was used to infer the rate of binary black hole and neutron star mergers, the mass distribution for both stellar objects, and the distribution of black hole spins. The mass distribution for neutron stars is found to extend from 1.2(+0.1-0.2) $M_\odot$ to 2(+0.3-0.3) $M_\odot$, though with outliers such as GW190814. The paper goes on to theorize that these systems are likely a small sub-population of neutron star-black hole systems, due to them being outliers in the binary black hole population. The authors do not explicitly support or reject the existence of a lower mass gap, despite previous papers on black holes being insistent that the gap exists. Instead it states that the existence of outliers and the general distribution of stellar masses calculated potentially indicates the lack of a lower mass gap entirely, but rather a continuous spectrum between neutron star and black hole masses.

Mandel and Farmer provide a comprehensive review of the formation channels for binary black hole systems [12]. In it, they discuss the methods of detecting black holes, those being gravitational waves and electromagnetic waves, the latter category including both x-ray binaries and other more specific methods such as microlensing. Later, they discuss formation methods for binary black holes. Binary star systems themselves cannot form binary black holes without the black holes either being too far apart to ever merge due to gravitational waves alone, or else being too close together so that one star engulfs the other during their red giant phase. The

authors discuss three main possibilities. One, the stars coming closer later in life, after the phase where the stars would have enveloped each other. Two, chemically homogeneous evolution, where a large star does not experience the typical expansion phase due to its chemical makeup, and therefore the stars can originate in a tight enough orbit to merge. Three, the stars did not originate within a binary at all, but rather the resulting black holes were moved together by the gravity of several other stellar objects. Of course, a combination of multiple of these causes is also possible. The paper goes on to discuss the methods for detecting different origins, their commonality throughout the universe, and the properties one can expect a binary black hole system formed through each method to exhibit.

In summary, the lower black hole mass gap was originally theorized to exist before the detection of gravitational waves, when x-ray binaries were the primary method of detecting black holes. However, it was not clear whether the gap was due to a fundamental property, or a flaw in the ability of stellar masses of that size to be detected. Gravitational waves, first detected in 2015, give scientists the chance to determine this once and for all, but although this fact has been acknowledged it has yet to be tested. Finally, if there are black holes within the mass gap they appear at a much lower frequency compared to larger mass black holes, which could potentially occur due to many reasons, one of the most likely being different common formation channels.

Methods

Event	M1	M2	χ_{eff}	p_{astro}
190704_104834	$7.090^{+3.642}_{-1.317}$	$3.921^{+0.845}_{-1.089}$	$0.132^{+0.150}_{-0.082}$	0.81 [14]
190814_211039	$24.483^{+0.816}_{-0.700}$	$2.722^{+0.057}_{-0.056}$	$0.0037^{+0.043}_{-0.038}$	>0.99 [4]
190821_124821	$9.441^{+4.903}_{-2.136}$	$4.477^{+1.213}_{-1.187}$	$-0.415^{+0.364}_{-0.174}$	0.60 [14]
190910_112807	$39.711^{+9.405}_{-3.460}$	$3.685^{+7.516}_{-0.448}$	$-0.650^{+0.825}_{-0.241}$	>0.99 [5]
190920_113516	$7.779^{+8.032}_{-1.981}$	$3.163^{+0.971}_{-0.280}$	$0.598^{+0.145}_{-0.136}$.	0.57 [14]
200210_092254	$29.225^{+6.321}_{-3.280}$	$3.3324^{+0.310}_{-0.384}$	$0.0346^{+0.173}_{-0.124}$	0.50 [3]
230529_181500	$3.997^{+0.590}_{-0.736}$	$1.427^{+0.279}_{-0.143}$	$-0.0493^{+0.081}_{-0.105}$	> 0.94 [16]

Table 1: This table shows all events used as data, excluding events with an astronomical probability under 0.5

I accessed the data from the Gravitational Wave Open Science Center [13]. I then downloaded and accessed the posterior samples for all events where one or both masses were between 2.5 $M_{\odot}$ and 5 $M_{\odot}$. I made histograms for each mass for each event, to determine with what certainty the mass was within this range, and therefore within the theorized mass gap. Figure 2.1 shows an example of these histograms. Events without one mass within the range

given two degrees of freedom were discarded. I then created a histogram of the chi effective for each remaining event. The chi effective is a number that tells us how aligned the spins of two black holes are. This is shown in figures 2.2 and 2.3. Next, I created a histogram averaging the chi effective for every remaining event. The chi effective's were weighted based on their probability of astronomical origin and the certainty they were within the appropriate mass range. These events are all displayed in table 1.
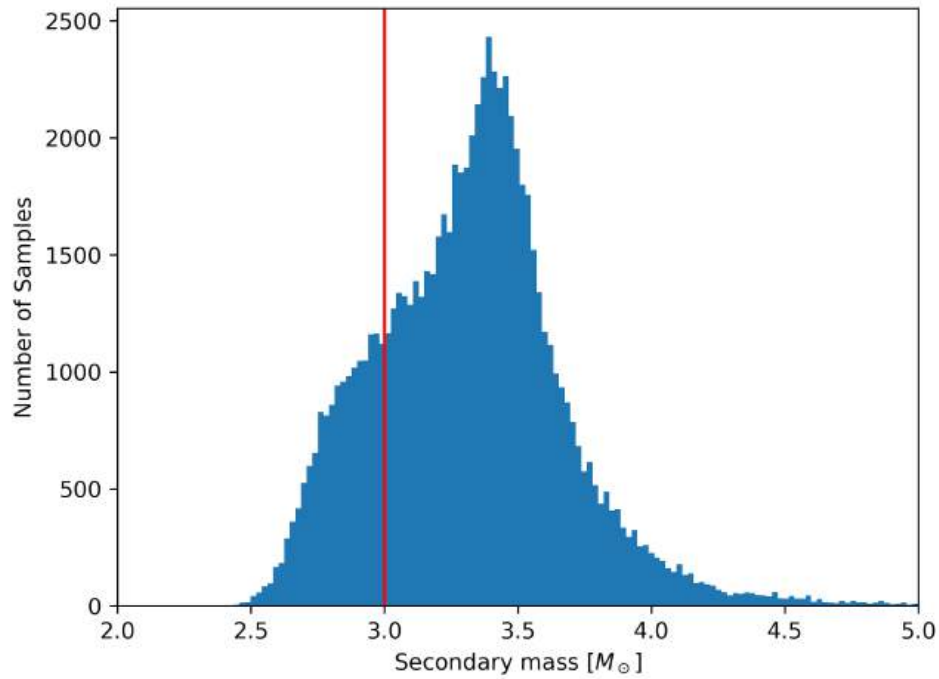
Figure 2.1 shows a histogram of probability distribution of the secondary mass from GW200210_092254, with a line drawn at 3 $M_{\odot}$ (the start of the 'lower black hole mass gap')
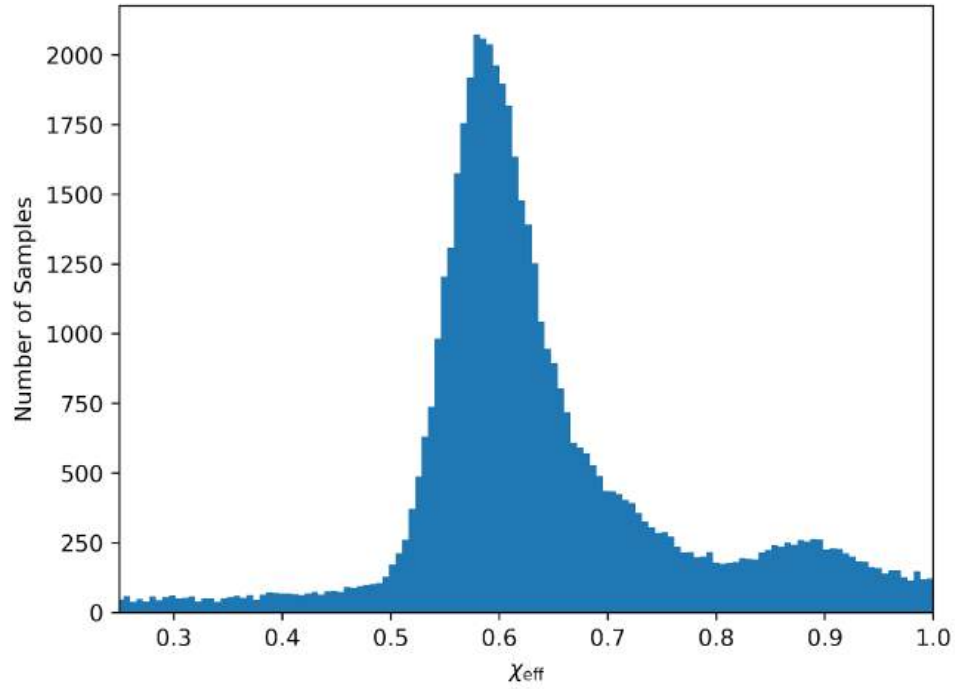
Figure 2.2 shows the chi effective distribution for GW190920_113516, as an example of an uneven distribution that does not align with the average value of 0.0104.
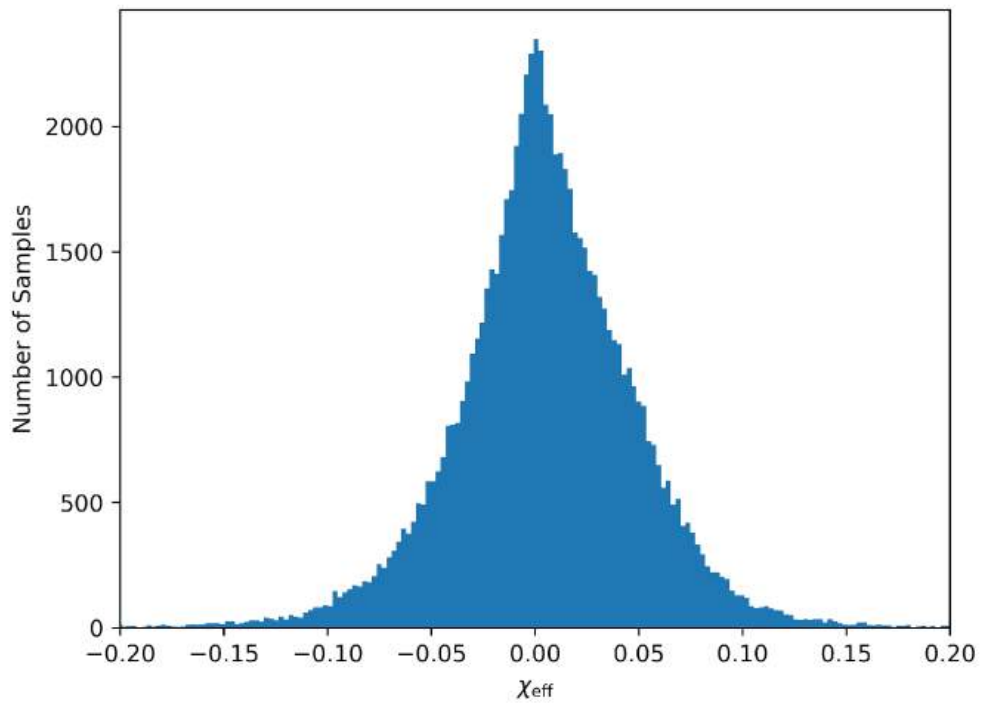
Figure 2.3 shows the chi effective distribution for GW190814_211039, as an example of an even distribution that aligns with the average value of 0.0104.

Results

        The events determined to contain masses within the range were mostly binary black hole mergers, although GW200210_092254 could also be either a binary black hole merger or a neutron star black hole merger, and GW230529_181500 could be either a binary neutron star merger or a neutron star black hole merger. The median chi effective from all events is $0.019^{+0.089}_{-0.074}$, as shown in figure 3.1.
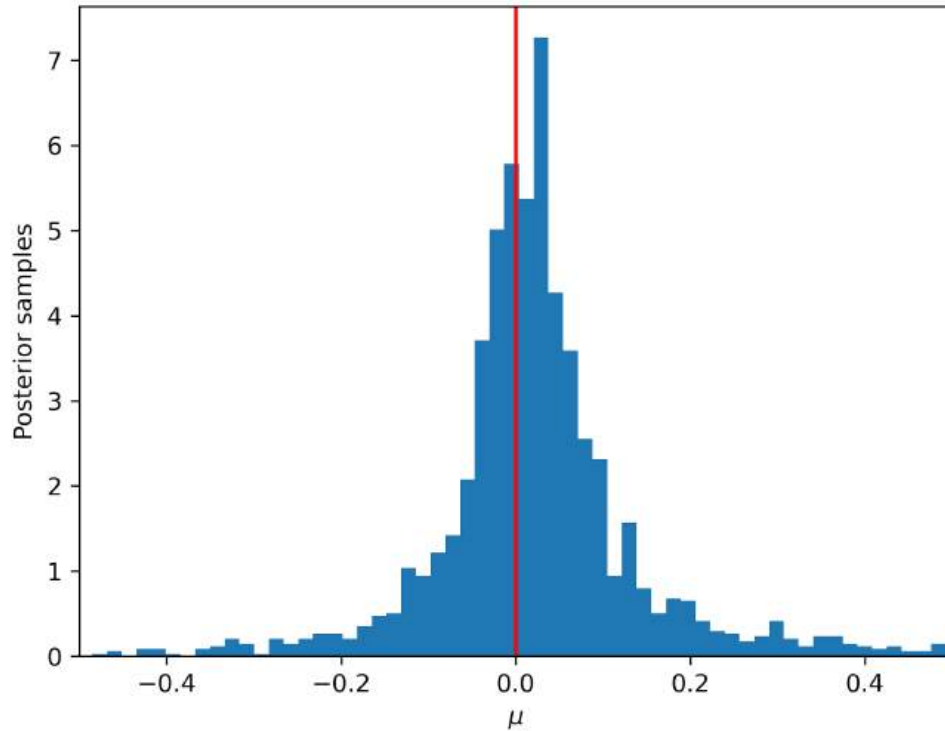
Figure 3.1 shows a histogram of the average chi effective values for all 7 events in the data set.

Discussion

        This study found eight gravitational waves with masses falling within the limits of 2.5 $M_\odot$ and 5 $M_\odot$ with at least 50% certainty. These masses have an average chi effective near 0. The chi effect ranges from -1 to 1, with values at either end representing spins that are perfectly aligned (parallel) and 0 representing spins that are entirely perpendicular. The existence of stellar masses within this range contradicts the existence of the previously theorized lower black hole mass gap. The average chi effective being near zero means that the spins of the black holes and neutron stars in these binary systems are on average perpendicular. In x-ray binary systems, the method for finding binary stellar mass systems before gravitational waves, the masses had formed from a binary star system, and therefore had parallel or near parallel spins, leading to a chi effective near positive or negative one. On the other hand, binary black hole and neutron star systems formed from star clusters- that is, closely grouped stars not part of one system, that

independently became black holes/neutron stars and became a binary system afterwards due to the gravitational pull of their surrounding stars- are much more random, and often have very unaligned rotations closer to zero. The chi effect of these systems within the previously theorized mass gap being near zero itself provides an explanation for why these systems were never found with x-ray binaries. For some reason, binary black holes and neutron stars at this in-between mass near exclusively form in star clusters, rather than x-ray binaries, and therefore were invisible to any detection methods relying on electromagnetic waves. Future research might be able to address this difference in black hole formation.

While this study identifies black holes in the mass gap, it is not without limitations. The primary limitation of this study was the size of the data set. Only 218 verified gravitational waves have been detected [16], and that pool is further limited by restricting the relevant masses. The data set would have been four times bigger at least in an ideal study, and the presence of only eight sets of data leaves room for error and uncertainty. A general pattern can be observed and analyzed, but it could be thrown off by the discovery of new data in LIGO observing round 4 and onward.

Now that this information- the lack of a lower black hole mass gap, and the formation patterns behind its members- has been discovered, the next step is the why. Why do black holes of this size seem to only form from star clusters? What about them prevents them from being seen in x-ray binaries, when no other range of masses seems to have this problem? The other work left to do is to continuously update this study with new data, as new gravitational waves fitting this pattern are discovered and have their data released.

Works Cited

[1] Abbott B.P., Abbott R., Abbott T.D., Abernathy M.~R., Acernese F., Ackley K., Adams C., et al., Observation of Gravitational Waves from a Binary Black Hole Merger. PhRvL, 116, 061102. 2016. doi:10.1103/PhysRevLett.116.061102

[2] Abbott R., Abbott T.~D., Acernese F., Ackley K., Adams C., Adhikari N., Adhikari R.~X., et al., Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. PhRvX, 13, 011048. 2023. doi:10.1103/PhysRevX.13.011048

[3] Abbott R., Abbott T.~D., Acernese F., Ackley K., Adams C., Adhikari N., Adhikari R.~X., et al., GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run. PhRvX, 13, 041039. 2023. doi:10.1103/PhysRevX.13.041039

[4] Abbott, R., Abbott, T Abraham, S. et al., GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object, Astrophys. J. Lett. 896, L44. 2020.

[5] Abbott, R. Abbott,T. Abraham, S. et al., GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run, Phys. Rev. X 11, 021053. 2021.

[6] Bailyn, C. D., Jain, R. K., Coppi, P., & Orosz, J. A. The Mass Distribution of Stellar Black Holes ApJ, 499, 367, 1998. doi:10.1086/305614

[7] Casares, J., Jonker, P.G. Mass Measurements of Stellar and Intermediate-Mass Black Holes. Space Sci Rev 183, 223–252 (2014). https://doi.org/10.1007/s11214-013-0030-6

[8] Clifford, Nick. Long-Term Timing of Pulsars in NGC 6440: An Updated Mass Limit of Millisecond Pulsar J1748-2021B, 2019. https://doi.org/10.18130/v3-bfjj-w888. doi.org/10.18130/v3-bfjj-w888.

[9] Farr W.~M., Sravan N., Cantrell A., Kreidberg L., Bailyn C.D., Mandel I., Kalogera V., The Mass Distribution of Stellar-mass Black Holes. ApJ, 741, 103. 2011. doi:10.1088/0004-637X/741/2/103

[10] Fryer, C. L., & Kalogera, V. Theoretical Black Hole Mass Distributions. ApJ, 554, 548, 2001. https://iopscience.iop.org/article/10.1086/321359

[11] Kalogera, V., & Baym, G. The Maximum Mass of a Neutron Star. ApJ, 470, L61. 1996. https://iopscience.iop.org/article/10.1086/310296

[12] Mandel I., Farmer A., Merging stellar-mass binary black holes PhR, 955, 1. 2022. doi:10.1016/j.physrep.2022.01.003

[13] "O4A Open Data Release." GW Open Science Center, gwosc.org/. Accessed 13 Sept. 2025.

[14] Olsen, S., Venumadhav, T., Mushkin, J., Roulet, J. Zackay, B., Zaldarriaga, M. 2022. New binary black hole mergers in the LIGO-Virgo O3a data. Phys. Rev. D. 106(4). 2022. https://journals.aps.org/prd/abstract/10.1103/PhysRevD.106.043009

[15] Ozel F., Psaltis D., Narayan R., McClintock J.E., The Black Hole Mass Distribution in the Galaxy. 2010, ApJ, 725, 1918. doi:10.1088/0004-   637X/725/2/1918

[16] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, Abac   A.~G., Abouelfettouh I., Acernese F., Ackley K., et al., GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run arXiv:2508.18082. 2025. doi:10.48550/arXiv.2508.18082

The Impacts of Indoor School Lighting on Circadian Rhythms and Overall Health By Carolyn Xiao

Indoor lighting in schools has been shown to disrupt circadian rhythms in humans, which is associated with negative health consequences such as cancer and obesity. Circadian rhythm disruptions due to the flaws of indoor lighting are detrimental to overall health. The circadian rhythm is the body's instinctive 24-hour clock; over time, this internal clock has evolved to sync with the external environment. Circadian rhythms are primarily organized by how light affects a brain region called the suprachiasmatic nucleus (SCN). Changes in brightness, spectral distribution, and the timing of light all have different impacts on the body. As a result, circadian rhythm disruptions are a common issue, especially among adolescents, who are more susceptible to irregular sleep-wake cycles. This is a major issue because the circadian rhythm regulates the body's hormone release cycle, metabolism, and body temperature. When these processes are disrupted, especially for a prolonged period, physical health may be endangered by obesity, high blood pressure, and cancer. To minimize circadian disruptions, a study on how different types of indoor lighting at school (LED, fluorescent, and natural light) affect the circadian rhythms and health of people would be helpful.

Light is first detected by the retina in the eye; the photoreceptors then relay the information to the suprachiasmatic nuclei (SCN), which is the masterclock of the body (Fisk). Light is crucial for a person's circadian rhythm to be synced with the environment; however, electricity offers unnatural sources of light that misaligns the body clock (Suni). In order to best sync with the environment, natural light has to be prioritized during the day. Specifically, natural light is beneficial in the morning since it encourages the body to feel alert throughout the day and tired at night (Patel). This is because exposure to natural light stimulates the production of cortisol - a hormone that promotes energy. Natural light also suppresses melatonin. Additionally, exposure to sunlight has major health benefits. During the peak of the COVID-10 pandemic, those who spend around 2 hours outside daily or in a bright room had experienced better sleep and less anxiety (Marshall). Exposure to natural light stimulates the brain's production of serotonin, which boosts overall mood and happiness. This is also a result of why bipolar depressed patients who were not exposed to natural light had a 3 day longer hospitalization period compared to those who were (Blumes).

Additionally, LED and fluorescent lights are helpful replacements for natural light, but they don't function as efficiently or effectively as natural sunlight does. For example, many patients who struggle with Seasonal Affective Disorder (SAD) are treated with bright light therapy (BLT). The treatment is carried through therapy lamps that function from illuminance levels 7000 to 10,000 lux; lux is a unit that measures how much visible light is received by the human eye (Blumes). For reference, full daylight without direct sunlight would have an illuminance level anywhere between 32,000 to 130,000 lux. The BLT lamp treatment plan has been recorded as successful; however, constant 1 hour morning walks in the sun are just as effective. This suggests that although BLT serves as a strong alternative, it does not fully

replicate the benefits of natural sunlight. Another source claims that fluorescent lights are damaging to the body clock: circadian rhythm patterns can be disrupted by fluorescent lights, leading to reduced melatonin and poor sleep (Make Great Light). By suppressing signals that promote alertness, melatonin, a hormone, encourages sleep and helps regulate the sleep-wake cycle. When large amounts of artificial blue light is present at night, the SCN doesn't receive the correct signals to initiate the release of melatonin. If the melatonin levels stay low, the natural sleep-wake cycle would be disrupted, which affects sleep quality and health. Finally, an experiment conducted proves the negative effects of LED lights, specifically at night (Jo). At 24:00, the average melatonin level during LED light exposure was drastically lower than under dim light. LED exposure had a 24.8% melatonin suppression compared to dim light, suggesting that LED lighting disrupts melatonin production at night. In conclusion, although artificial light acts as possible replacements for natural light, it cannot fully replicate the benefits, which disrupts sleep quality and circadian health.

The inquiry approach for this project was to use comparative analysis and case study, and the study focused on how different types of indoor lighting at school affected the circadian rhythms and overall health of people. Data was collected through conducting surveys and interviews with teachers who experienced consistent indoor lighting throughout the day. To understand the data, both qualitative and quantitative methods were used, including interviews to gather detailed qualitative responses and general interview questions for broader quantitative insights. The formula for this project involved using both primary and secondary sources, as data was collected from interview questions as well as existing research on how lighting affects circadian health. Therefore, a descriptive approach was used because observational data was gathered from teachers without making experimental changes to the environment, such as lighting alterations. The data was analyzed using thematic analysis to highlight common patterns in narratives and possibly statistical analysis as well. Statistical analysis ended up not being conducted. This research project specifically focused on group study and case study, which provided valuable information. Finally, the goal for this process was that the results would help identify themes between participants and light exposure, providing insights on how indoor lighting might have affected circadian rhythms and overall health. After initially gathering the qualitative responses from a total of 12 participants through google forms. After initially collecting qualitative and quantitative responses from a total of 12 participants through Google Forms, two participants were selected for follow-up interviews to gather more information. These two participants were the focus for the specific case studies in this project. 15 minute interviews were conducted in the following 2 weeks. After, Participant 1 and Participant 2 voluntarily discussed their experiences in more depth and expressed interest in collecting more data. Many methods were considered. The first idea was to have the participants recording their meals and food intake. This was considered because food intake and timing is a powerful metric for circadian rhythms. However, this method was too restrictive and time consuming for the participants. The next possible experiment was recommended by a mentor; it was suggested that data could be taken from a device such as an apple watch or a fitbit. These devices provided

valuable information such as heart rate, sleep duration, sleep quality, ect. Participant 1 had a fitbit, so data was taken from that over a 1 week period. Participant 2 consented to using a sleep tracking app called Sleep Cycle. This app gave crucial information just like the fitbit. Data was recorded and analyzed from both participants. Note that this research was conducted at a student level and does not reflect a professional medical diagnosis or advice. All participants consented prior to taking the surveys, and the questions were verified through the IRB process.

      Participant 1's most energetic time of day is in the morning. However, they rely heavily on caffeine to feel alert in the morning, which shows that although their circadian rhythm is well-aligned with their wake-up time, their natural alertness is far from optimal. Participant 1's body is dependent on external stimulation, which may indicate lowered cortisol levels. Cortisol is a hormone that induces feelings of alertness by increasing energy levels (Cleveland Clinic). If cortisol levels are low when waking, Participant 1 won't feel energized during the morning. This response could be impacted by late-night device exposure, since blue light suppresses melatonin- melatonin is a hormone that regulates sleep by signaling to the body that it is time to rest (Sleep Foundation). Melatonin suppression inhibits the natural cortisol increase in the morning. Incorrect cortisol release can also be influenced by a misaligned circadian clock. Another possible reason for lowered natural alertness could be disruptions during a stage of sleep called the Slow-Wave Sleep (SWS). SWS is the deepest and most restorative part of the Non-Rapid Eye Movement sleep (NREM), occurring earlier during the night (Newsom). SWS is characterized by lowered brain activity, body temperature, and heart rate. During this stage, high-amplitude delta waves (0.5-4 Hz) are present in the brain. These particular waves display a state of deep sleep and lowered cognitive activity. However, even though SWS occurs during the first half of the night, it can still impact morning alertness. SWS controls adenosine levels, which are responsible for sleep pressure and the urge to sleep- elevated adenosine levels in the body cause drowsiness. During deep sleep (SWS), the brain clears excess adenosine, reducing the sleep pressure (Newsom). This also promotes wakefulness in the morning. If SWS is disrupted, adenosine clearance may be incomplete, leading to residual sleepiness after waking up. Disruptions of the SWS stage may contribute to Participant 1's dependence on caffeine for alertness during the morning. Next, Participant 1 feels most tired after lunch around 2:00 PM and again in the evening at 8:00 PM. The afternoon dip in energy is likely a result of a high carbohydrate lunch, causing abrupt blood sugar changes. When blood sugar (glucose) levels rise, insulin is released to utilize the glucose for energy. When too much insulin is released, blood sugar is lowered fast. This process is called postprandial hypoglycemia, and is responsible for the sudden drop in energy (Mayo Clinic). The higher the intake of carbohydrates, the stronger the blood sugar crash, leading Participant 1 to feel more tired. Participant 1 has strong sleep onset. They do not take melatonin prior and are exposed to blue light directly before sleeping. However, Participant 1 takes around 10-15 minutes to fall asleep, and rates their overall sleep satisfaction a 4/5. Next, Participant 1 is exposed to at least 5 hours of fluorescent and LED light per day, and experiences a 7 hour screen time average in total. Occasionally, they experience blurry vision after looking at a screen, suggesting digital eye strain. Constant screen usage may affect sleep depth later that

day, which could contribute to the lowered morning energy Participant 1 experiences. Participant 1 has a very consistent eating schedule with breakfast at 6:30 AM, lunch at 12:00 PM, and dinner at 6:30 PM. Eating schedules are crucial to regulating circadian rhythms (Wehrens). The human body is composed not only by the central clock in the brain (SCN), but by many peripheral clocks - biological clocks existing in organs and muscles. Peripheral clocks are especially sensitive to food time and intake, and they regulate processes like digestion, energy consumption, and metabolism. As a result, meal timing is a powerful circadian metric. Participant 1's energy fluctuations, morning caffeine dependence, and vision blur may indicate disruptions in SWS, cortisol levels, and postprandial hypoglycemia (Wehrens). Further investigation into Participant 1's caffeine intake, (blue) light exposure, and meal composition may help identify specific factors affecting their circadian rhythms.

Next, Participants 2-10 all spend their school hours in the same building, therefore they receive similar conditions of indoor light and workload. However, their energy levels, sleep patterns, and overall condition vary greatly, showing that individual differences in lifestyle are more impactful than the school environment. 3 of the 10 participants said that the fluorescent lighting is way too harsh, often glaring off of screens and the whiteboard. From a health aspect, prolonged exposure to intense artificial light contributes to eye pain, fatigue, and can even reduce alertness (Chepesiuk). The other 7 participants were not fond of artificial light, but accustomed to it. All of the ten participants feel the most energetic during the morning, specifically mid-morning. This is common among the participants that teach an early morning class. It can be concluded that teaching earlier during the day, instead of having a prep period, increases adrenaline from student-teacher interactions. The 2 participants that feel energetic later (11:00 am - 1:00 pm), likely experience disrupted cortisol rhythms, as explained for Participant 1. One common trend among most of the 10 participants was an afternoon slump around 2:30 PM. A possible reason for this could be fatigue from teaching courses earlier during the day. 7 out of the 11 participants stated that they feel the most fatigued around 2:00 PM to 3:00 PM. 2 of the participants claim this is because adrenaline from teaching all day subsides, since many of the participants have been teaching nonstop at this point. However, this afternoon dip in energy is a frequent occurrence in circadian biology as well. The process is run by naturally occurring circadian rhythms and homeostatic sleep pressure. Sleep pressure or the desire to sleep is driven by adenosine accumulation in the brain (Huang). By afternoon, many people experience a sensitivity to adenosine because this is when the accumulation reaches a moderate amount. Additionally, participants experience excess sleepiness when this adenosine increase is overlapped with having a high-carbohydrate lunch prior (Wehrens). It is also important to note that "morning people" or those who start their day earlier experience more fatigue in the afternoon. This is because their circadian rhythm is trained to promote earlier wakefulness, driving their peak earlier in the day, which declines sooner. All of the participants wake up relatively early, making them all more susceptible to this afternoon dip, as shown in the data. Next, many participants are especially concerned about screen exposure and blue light, with some experiencing eye strain. The 10 participants have a wide range of screen exposure time,

with the least being 2 hours and the most being 9 hours. This is similar to the time spent under indoor lighting.

Participants 2-11 Daily Exposure time to Classroom Lighting
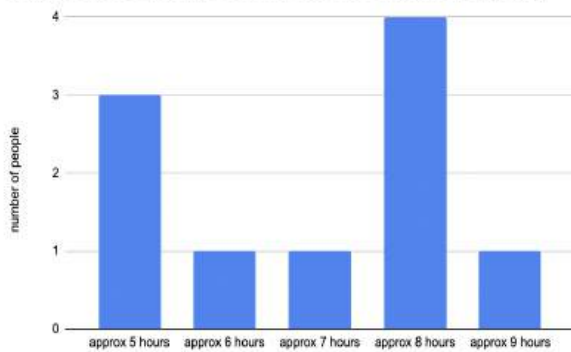
Participants 2-11 Daily Screen Exposure
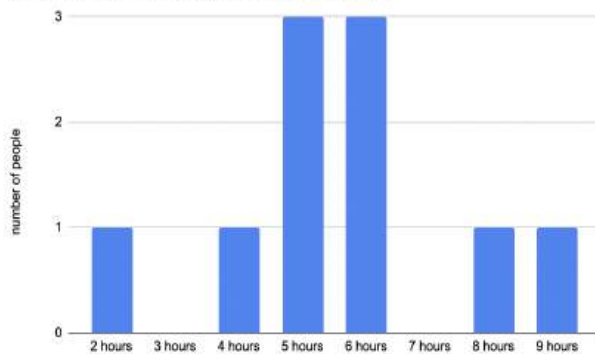
Have you experienced changes in your sleep schedule during the school year that you think might be linked to your time spent under classroom lighting?

11 responses

- Yes
- No
- Maybe

36.4%
63.6%
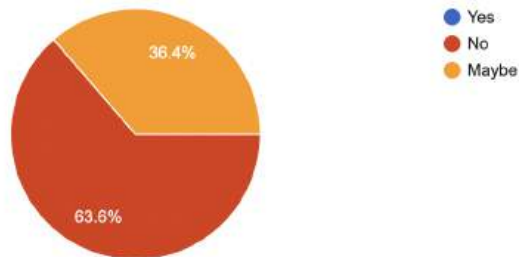
On a scale of 1-5, what would you rate your satisfaction with your sleep? (5 being high satisfaction and 1 being low)

11 responses

- 1
- 2
- 3
- 4
- 5

54.5%
45.5%
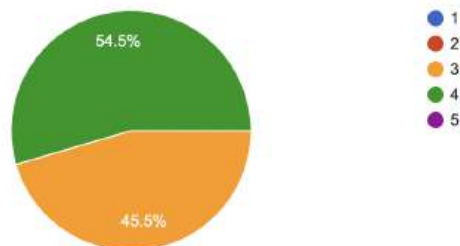
Unlike all of the previous participants, Participant 11 has more flexibility with changing locations throughout the day. This inconsistent exposure suggests that their circadian rhythm is heavily influenced by environmental changes rather than prolonged indoor lighting conditions. However, they are exposed to a consistent classroom environment at least 4 hours each day. Participant 11 has florescent lights and natural light available in their classroom. But, they don't frequently use the overhead fluorescent lights, and prefer to only rely on natural light. While natural light is recognized as the best source of lighting for one's circadian rhythms, there are some consequences with relying on it in an indoor setting (Blume). In the morning, the intensity of sunlight is relatively low. This is especially the case for classrooms because windows filter out some wavelengths of light (Siraji). The sun is lower in the sky before noon, meaning light has more to travel to reach Earth's atmosphere. This ends up scattering shorter wavelengths, such as blue or violet, reducing the strength of natural light we receive. Additionally, standard window glass is able to block most of the UV spectrum, which is below 400 nm. Therefore, it is also able to affect the presence of blue light (400-450 nm). This means that even when a classroom is fully lit by sunlight, it is possible that natural light indoors is less effective than fluorescent light (400-700 nm). Next, similar to the other participants, Participant 11 feels the most energetic at the start of the day. They feel the least energetic during the afternoon. This participant also mentioned that they feel tired during the day when their kids wake them up in the middle of the night. When someone is woken up 4 hours after falling asleep, which is typically the middle of the night, they are likely to be in the middle of their third or fourth sleep cycle. This stage is the deep sleep phase (NREM). NREM is the most restorative stage of sleep, which is crucial for memory processing, the immune system, and physical recovery (Cleveland Clinic). Interruptions can leave Participant 11 feelings fatigued up until the next day. If this is a frequent matter, their overall health and cognitive function will be affected.

In order to strengthen the analysis, further data collection was done with Participants 1 and 2. Sleep data was taken from March 13 to March 19. Among all of the nights, March 13 gives the most comprehensible data, making it the best reference point for comparison. On the night of March 13, Participant 1 recorded 6 hours and 56 minutes of sleep, with a sleep score of 76, a resting heart rate of 58 bpm, and movement occurring during 7% of the sleep period. Overall, participant 1 experienced a fairly consistent resting heart rate, ranging from 54 to 59 bpm, with the highest recorded on March 17 (59 bpm). The sleep score was not as consistent, fluctuating from 76 to 85. Sleep score is affected by various factors, such as sleep duration, sleep cycles (light, deep, REM), the number of awakenings, and movement. It is likely that the higher movement (7%) experienced on March 13 contributed to the lower sleep score of 76. Participant 1's sleep duration varied from 6 hours 10 minutes to 7 hours 61 minutes, which averages close to 7 hours per night. While this participant maintained a regular sleep schedule, the restorative quality varied based on other factors.

This sleep data gathered supports the analysis of Participant 1 gathered in previous paragraphs. It is difficult to determine when the 7% of movement during sleep occurred, but

inferences can be drawn from how movement affects sleep stages. Movement during the SWS stage of sleep is rare but very disruptive. Mentioned before, SWS is the most restorative stage of sleep, characterized by low metabolic activity and high neural synchronization.

Additionally, heart rate is a strong biomarker for sleep recovery and quality (Nava). Participant 1's resting sleep heart rate ranges from 54 bpm to 59 bpm. This is preferred because lower sleep heart rates indicate reduced nervous system activity. It also indicates higher parasympathetic dominance. Parasympathetic dominance refers to a state where the parasympathetic nervous system (PNS) is more active than the sympathetic nervous system (SNS). PNS is a stage known for lowered heart rate, high levels of digestion, and is preferred because it allows the body to be in a restorative state (ScienceDirect). This is highly beneficial for tissue repair and memory consolidation. From this data, it can be concluded that Participant 1's cardiovascular system functions efficiently during periods of sleep.

The opposite is true when circumstances are the opposite. On the night of March 17 for example, the heart rate was the highest it has been during this recorded period. Participant 1 experienced a heart rate of 59 bpm. As expected, the sleep score dropped down to 79, which is notably lower than previous nights. This drop may indicate wakes throughout the night, higher caffeine intake, or stress.

The additional data for Participant 2 was collected through an app called Sleep Cycle. Sleep Cycle is a commonly used app; however, it is not a clinical-grade device. This means that the Sleep Cycle is unable to track the exact stages of sleep like how EEG devices do. EEG or electroencephalogram, is a clinical tool used to record the brain's electrical activity. Instead, Sleep Cycle uses noises and movement the user makes throughout the night to approximate the different stages of sleep: light, deep, awake, or REM. It is important to note that following analysis conducted on Participant 2 should be interpreted with caution, as physiological measurements were not used. This analysis is not conducted by a licensed medical professional, and is not a diagnosis in any form.

Participant 2's sleep data was recorded from March 20 to March 23. This participant's time in bed ranges from 5 to 8 hours each night, with nearly an average of 6 hours. However, the actual sleep time is only around 5.4 hours each night. The breakdown shows that the duration of deep sleep is extremely low with 3 minutes on the 20th, 10 minutes on the 21st, and 34 minutes on the 22nd, and 10 minutes on the 23rd. Adults are recommended to experience at least 60 minutes of deep sleep each night. It is clinically improbable for the average adult to get this little deep sleep, so it is likely that this was an error with the app's tracking. Deep sleep, also referred to previously as slow-wave sleep (SWS), is essential for physical recovery. This is the stage where the body strengthens the immune system and repairs muscle fibers. Receiving 34 minutes or less of deep sleep each night leads to major health issues. Participant 2 would end up feeling very sleepy during the daytime, since deep sleep is essential for reducing sleep pressure. Over

time, the lack of deep sleep leads to increased risks of type 2 diabetes, obesity, and various cardiovascular diseases. Again, it is unlikely that Participant 2 experienced this little deep sleep.

Another common factor was the snoring Participant 2 experienced. On March 21 and March 22, participant 2 experienced around 3 hours of snoring. This indicates disrupted breathing or airway resistance. Both of these factors end up reducing the amount of oxygen the body receives, and may be a symptom of sleep apnea. Snoring occurs when the soft tissues in the throat vibrate (Mayo Clinic). This occurs when the airway is blocked. There are many causes for airway obstruction, the most common being nasal congestion, enlarged tonsils, or being tongue-tied (medically known as ankyloglossia). Ankyloglossia is a condition where the band of tissue connecting the floor of the mouth to the bottom of the tongue is unusually short. This is a major cause for snoring, as can contribute to sleep apnea.

Next, Participant 2 had consistent bedtimes - around 10:30 for the days recorded. The wake times were also consistent, leading to a stable regularity score. On March 21 and March 22, Participant 2 experienced regularity scores of 84% and 88%, respectively. This shows that they have a consistent and healthy circadian alignment. However, the difference in the sleep regularity and the sleep quality recorded above suggests that bedtime regularity isn't the main factor in determining the overall experience of sleep. Additionally, it is important to note that Participant 2 reported that total sleep duration has a greater impact on their performance and how rested they feel, whereas the quality of their sleep.

Night of March 19 - March 20

Night of March 20 - March 21

11:43 PM   1   2   3   4   5:41 AM

- Awake   0h 30m
- Light   5h
- Dream   0h 25m
- Deep   0h 03m

Show more

Night of March 21 - March 22

10:36 PM   12   1   2   3   4   5   6   7:45 AM

- Awake   1h 24m
- Light   6h
- Dream   1h 12m
- Deep   0h 34m

Show more

Night of March 22 - 23
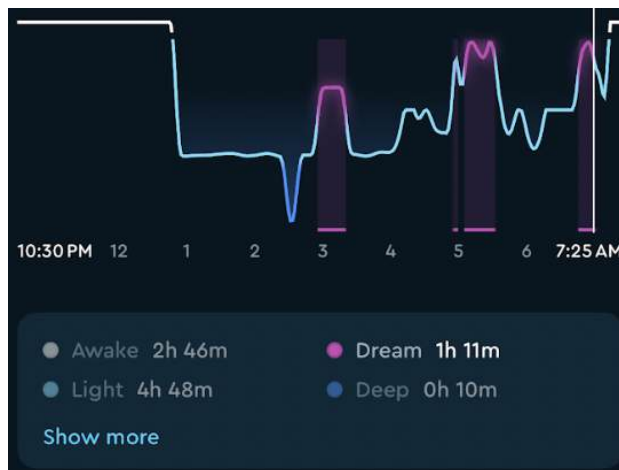
Snore Data:

This study analyzed the effects of indoor school lighting on circadian rhythms and overall health. Several conclusions can be drawn from the analysis of qualitative and subjective participant response and the objective sleep data recorded. The most important conclusion is that although indoor lighting factors play a role in influencing circadian rhythms, individual lifestyle factors such as habits, meals, and light exposure outside school have a bigger impact. This is especially visible with the qualitative analysis done with Participants 2 through 11, who all experienced a similar school environment. Even though they shared many work factors, the reported energy levels, sleep, and health varied. Some participants felt more energized in the

mornings, and others did not. However, many of these 10 participants experienced an afternoon crash. Additionally, most of these participants shared that they disliked the harshness of the indoor lighting and believed it to directly affect their health. Once again, this proves that individual lifestyle factors and biological differences affect circadian rhythms more than environmental lighting. Overall, this specific trial done with participants 2 through 11 shows that indoor lighting is only a minor contributing factor to circadian health.

However, the specific case studies conducted on Participant 1 and Participant 2 suggest that the school lighting conditions may play a role in influencing sleep and circadian patterns. Participant 1 is very content with the high amount of natural light they experience in their school environment. In contrast, Participant 2 is mainly exposed to harsh fluorescent indoor light and reported dissatisfaction with their environment. It can be possible that Participant 1's adequate school environment is what leads them to have a low sleeping heart rate, high sleep satisfaction, and a fairly high sleep score (76-85). Likewise, Participant 2's bad lighting conditions can be related to their low deep sleep time and frequent snoring. Additionally, Participant 2 experienced a sleep latency as high as 2 hours and 28 minutes compared to Participant 1, who experienced around 15 minutes. These results suggest that natural lighting availability contributes more to circadian health than indoor artificial lighting alone. In conclusion, this research shows that while individual lifestyle and biological factors have the biggest impact on circadian health across a large group, the quality and accessibility to natural light can help individuals regulate their circadian rhythms.

The main and most prominent finding is the importance of natural lighting and lifestyle/biological factors. From the data collected, these elements have the largest effect on circadian health. The impact of natural lighting was shown in the case studies of Participant 1 and Participant 2. The importance of individual and biological factors was shown in the group study (Participants 2 through 11). Since sleep is a universal experience and stuff, the implications of this research should matter to everyone. However, this specific study is meant to inform school administrators, educators, and adolescents (10-19 years old). Adolescents are especially vulnerable to circadian misalignment because of biological shifts. If adult participants were affected in this study, it can be concluded that students in the same environment would be affected even more. The findings of this study aligns with previous research conducted on the importance of natural light. Previously cited studies (Jo) in the introduction support the idea of artificial light not replicating the unique benefits of natural light. My findings agree with the past findings. They both show that common use of artificial light has detrimental effects on sleep quality. Similar to the Cleveland Clinic, these findings prove that individual habits like meal timing and caffeine use greatly influence circadian rhythms. One major limitation of this project would be the lack of non-clinical tools. Qualitative responses and non-clinical apps like the Sleep Cycle app are not always accurate. Qualitative responses are also subject to bias and are based on perception. If further research were to be done, more measurable data collection methods could be included. It would be more accurate to use a clinical sleep tracker, like an EEG for example.

The school environment itself could also be modified in order to allow controlled testing. This would help isolate the various factors and give fewer variables when testing how circadian rhythms are affected.

Works Cited

Moore-Ede, Martin, et al. "Lights Should Support Circadian Rhythms: Evidence-Based
Scientific Consensus." Frontiers in Photonics, vol. 4, 2023,
https://doi.org/10.3389/fphot.2023.1272934. Accessed 3 Oct. 2023.

Global Circadian Rhythm Lighting Market – Industry Trends and Forecast to 2030. Data Bridge
Market Research,
https://www.databridgemarketresearch.com/reports/global-circadian-rhythm-lighting-mar
ket. Accessed 24 Oct. 2024.

Esaki, Tsunetaka, et al. "Impact of blue-enriched light on sleep and circadian rhythms: A
systematic review of randomized controlled trials." International Journal of
Environmental Research and Public Health, vol. 17, no. 15, 2020, Article 5291,
https://doi.org/10.3390/ijerph17155291. Accessed 24 Oct. 2024.

Figueiro, Mariana G., et al. "Circadian-effective light and its impact on alertness in office
workers." Nature and Science of Sleep, vol. 11, 2019, pp. 113-122,
https://doi.org/10.2147/NSS.S194743. Accessed 24 Oct. 2024.

Marshall, Ali. "How Morning Sunlight Exposure Can Benefit Your Health." Verywell Health, 16
May 2023, https://www.verywellhealth.com/morning-sunlight-exposure-3973908.
Accessed 24 Oct. 2024.

"Morning Light for Better Sleep." WebMD,
https://www.webmd.com/sleep-disorders/features/morning-light-better-sleep. Accessed
24 Oct. 2024.

Obayashi, Kenji, et al. "Exposure to Light at Night and the Risk of Depression in the Elderly."
Journal of Affective Disorders, vol. 226, 2018, pp. 172-179,
https://doi.org/10.1016/j.jad.2017.09.036. Accessed 24 Oct. 2024.

Effect of Fluorescent Lighting on Health. Make Great Light,
https://www.makegreatlight.com/lc/effect-of-fluorescent-lighting-on-health. Accessed 24
Oct. 2024.

Souman, J. L., et al. "Lighting in Offices: Visual and Non-Visual Effects of Lighting on Office
Workers." International Journal of Environmental Research and Public Health, vol. 18,
no. 2, 2021, Article 5378, https://doi.org/10.3390/ijerph18105378. Accessed 24 Oct.
2024

Randler, Christoph. "Morningness–Eveningness and Sleep–Wake Variables in Adolescents."
Sleep and Biological Rhythms, vol. 6, no. 4, 2008, pp. 223–229. PubMed Central,
https://doi.org/10.1111/j.1479-8425.2008.00372.x. Accessed 19 Nov. 2024.

Cleveland Clinic. "Cortisol: What It Does & How To Regulate Cortisol Levels." Cleveland
Clinic, 14 Sept. 2023, https://my.clevelandclinic.org/health/articles/22187-cortisol.
Accessed 26 Feb. 2025.

Sleep Foundation. "Melatonin: What It Is & How It Works." Sleep Foundation, 22 Dec. 2023,
https://www.sleepfoundation.org/melatonin. Accessed 26 Feb. 2025.

Newsom. "Slow-Wave Sleep: An Overview." Sleep Foundation,
https://www.sleepfoundation.org/stages-of-sleep/slow-wave-sleep. Accessed 26 Feb.
2025.

"Reactive Hypoglycemia: What Causes It?" Mayo Clinic, 14 Sept. 2023,
https://www.mayoclinic.org/diseases-conditions/diabetes/expert-answers/reactive-hypogl
ycemia/faq-20057778. Accessed 26 Feb. 2025.

Wehrens, Sophie M. T., et al. "Meal Timing Regulates the Human Circadian System." Current
Biology, vol. 27, no. 12, 2017, pp. 1768–1775.e3.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5483233/. Accessed 26 Feb. 2025.

Chepesiuk, Ron. "Missing the Dark: Health Effects of Light Pollution." Environmental Health
Perspectives, vol. 117, no. 1, Jan. 2009, pp. A20–A27.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2627884/. Accessed 26 Feb. 2025.

Huang, Zhi-Li, et al. "Adenosine A2A Receptor Activation Induces Sleep Through the
CaMKIIα–p38 Pathway in the Lateral Preoptic Area." Sleep Medicine, vol. 110, 2024, p.
103. https://www.sciencedirect.com/science/article/pii/S1389945724000534. Accessed
26 Feb. 2025.

Siraji, Mushfiqul Anwar, et al. "Effects of Daytime Exposure to Short-Wavelength-Enriched
White Light on Alertness and Cognitive Function Among Moderately Sleep-Restricted
University Students." Building and Environment, vol. 255, 2024, p. 110861.
[https://www.sciencedirect.com/science/article/pii/S0360132324000878](https://www.sci
encedirect.com/science/article/pii/S

"Sleep: What It Is, Why It's Important, Stages, REM & NREM." Cleveland Clinic, 19 June
2023, https://my.clevelandclinic.org/health/body/12148-sleep-basics. Accessed 26 Feb.
2025.

"Parasympathetic Nervous System." *ScienceDirect*, Elsevier,
www.sciencedirect.com/topics/neuroscience/parasympathetic-nervous-system. Accessed
28 Mar. 2025.

Mayo Clinic Staff. "Snoring." *Mayo Clinic*, Mayo Foundation for Medical Education and
Research, 24 Feb. 2024,
www.mayoclinic.org/diseases-conditions/snoring/symptoms-causes/syc-20377694.
Accessed 28 Mar. 2025.

Zisapel, Nava. "New Perspectives on the Role of Melatonin in Human Sleep, Circadian Rhythms
and Their Regulation." *British Journal of Pharmacology*, vol. 179, no. 10, 2022, pp.
1342–1356. *PubMed Central (PMC)*,
https://pmc.ncbi.nlm.nih.gov/articles/PMC8196961/. Accessed 28 Mar. 2025.

Oxford's Philosophy, Politics, and Economics By Arjun Rajkumar

What makes a leader truly effective: the sharpness of their logic, or their ability to reflect on the past? Since its establishment in 1920, Oxford's Philosophy, Politics, and Economics (PPE) degree has become an academic must for political figures. The degree itself has produced twenty-one presidents and prime ministers from nine nations. Designed as a response to major global conflicts, such as World War I, that had diminished classical learning, PPE sought to arm future politicians with a blend of normative reasoning, institutional analysis, and empirical understanding, all tailored for twentieth-century governance (Dr. Jonathan Anomaly). In many ways, the degree has been successful: It has nurtured generalists capable of managing and interpreting the law and managing economic systems efficiently (Needleman).

However, the question remains: Does technical versatility equate to balanced leadership, including moral and historical insight? Many critics of the PPE degree suggest that its graduates, despite their brilliant strategies and versatility, may lack the deep contextual literacy and moral humility that a historical degree could provide. History, unlike PPE, provides its students with an understanding of narratives, patterns, and the consequences of power. Out of the forty-seven presidents, six have majored in History. This list includes notable figures such as JFK and FDR, showcasing its diversity among important historical figures ("Famous History Majors | History at Illinois"). The degree offers a broad perspective on the fragility of institutions, the contingency of progress, and the consequences of repeating past mistakes, making it ideal for politicians who want to reflect on the past to inform their decisions ("What Does PPE Stand for and Why Study a PPE Degree?").

Although the study of history offers an equally vital foundation by cultivating deep contextual awareness, ethical reflection, and a nuanced understanding of how the past shapes the present, Oxford's PPE degree equips future leaders with the interdisciplinary clarity necessary for effective policymaking, making it a better choice for politicians ("What Does PPE Stand for and Why Study a PPE Degree?").

The Role of History in Political Education and Its Limits

For centuries, history was regarded as the main foundational discipline for politicians and rulers. In both Eastern and Western traditions, it was seen as a vital pathway to understand how societies function, how power was used, and how mistakes could be repeated. In Imperial China, for example, historical knowledge was a prerequisite for entry into civil service; it was believed that political and moral judgment could only be achieved through a deep understanding of the past dynasties and how they rose and fell (L.). As David Foster explains, "[In ancient China] scholars were expected to know history well, even with changes in time and dynasty, a study of the past allowed one, it was said, to 'cope with the myriad through the unvarying

principle'"(Foster; "Why Are History Majors Successful in Political and Public Policy Careers? | Department of History"). Therefore, wise rulers were the ones who used history as guidance for the present. These scholars were entrusted with running the government and, more importantly, with providing living examples of moral leadership for the larger society." This mindset continued into the modern world. American presidents like Franklin D. Roosevelt and John F. Kennedy, both of whom studied history, used their knowledge in their presidencies ("Famous History Majors | History at Illinois"). Roosevelt, a history student at Harvard, often gained inspiration from the American Civil War and the Progressive Era to frame the New Deal during the Great Depression, while Kennedy's speeches regularly invoked classical and historical parallels to lend moral authority to Cold War policy decisions (Burke).

However, while historical learning may improve public language and high-level judgment, it is less effective for the economic and analytical needs of twenty-first-century politics. History prioritizes narrative understanding over empirical modeling or policy design. It encourages retrospection, not necessarily intervention. As global governance is increasingly reliant on the economy and AI, history is slowly becoming out of touch due to the fact that this era of technology and politics has never happened before. Leaders must master the ability to synthesize moral reasoning, institutional logic, and quantitative analysis—skills not typically cultivated through historical study alone. For example, a leader managing inflation, regulating artificial intelligence, or negotiating trade agreements must do more than reflect on the past; they must understand fiscal theory, legal systems, and the political implications of economic decisions. History may suggest caution or precedent, but it rarely provides practical instruction.

Why PPE is a favorable choice for Politicians

Unlike history, which emphasizes retrospection and caution, Philosophy, Politics, and Economics (PPE) was designed to address the intellectual demands of modern governance (Wheaton College). Its structure reflects an understanding that leadership in the modern era of politics can not rely solely on narrative insight and classical thinking from the past. PPE was designed "as something radical" at Oxford in 1920, "in the aftermath of the Russian Revolution and the First World War," and as an alternative to the Great (Classics) (Beckett). Oxford had, according to its official history, become 'interested in the problems raised by political unrest in Europe and Asia and high unemployment in the UK,' aiming to prepare upcoming politicians and government officials for a more complex and institutionalized world (Beckett). Instead of focusing on the past, the degree trains students to think about the systems that define modern politics: ethics from philosophy, the ideologies from government, and the complexity of the economy. It is a curriculum that is not rooted in caution, but action (Spring Hill College). Oxford PPE began as something new (Beckett). In 1920, in the aftermath of the Russian Revolution and the First World War, the university was in a reforming phase (Beckett). Usually glacially slow to

change, Oxford became "interested in the problems raised by political unrest in Europe and Asia and high unemployment in the UK" (Beckett).

In today's increasingly complex world, the structure of PPE has become even more relevant. Today's leaders must navigate through global markets, quantitative data, and reconcile public interest, almost always under pressure. For instance, when evaluating legislation to expand on civilian surveillance, a politician must first consider the implications of said legislation—whether such measures infringe on principles of justice or violate a human's rights—drawing on philosophical reasoning. Next, they must consider the viability and long-term economic consequences of implementation. Finally, they must navigate the political landscape of the legislation: gaining support for the legislature, predicting public opinion, and strategizing on how to gain votes for the next passage.

However, the politician can't use historical knowledge in this scenario because legislation and scenarios like this have never occurred in history before. As the Duke University Political Science Department explains, " The original idea was, in part, to provide future civil servants in the United Kingdom with an opportunity to become generalists by exploring public policy through the different lenses of complementary disciplines." (Anomaly). By combining normative reasoning with institutional and economic analysis. This combination creates some of the best politicians the world has to offer.

Conclusion

For centuries, history has been considered to be an essential part of political leadership. For politicians around the world, it provided insight into power and the consequences of past mistakes. In Imperial China, historical knowledge was a must for civil service, viewed as a key portion of sound moral judgment. As David Foster notes, "a study of the past allowed one… to 'cope with the myriad through the unvarying principle.'" (Foster; "Why Are History Majors Successful in Political and Public Policy Careers? | Department of History"). Even in more modern times, leaders like Franklin D. Roosevelt and John F. Kennedy—both trained in history–used their knowledge of the past to frame their political careers. However, while history strengthens moral awareness and reflection on the past, it doesn't equip politicians with the skills to combat problems that are frequent in the modern world.

PPE, on the other hand, was created in Oxford during the year 1920 to prepare students specifically for those demands (Anomaly). The degree was developed as a response to war and revolution (Anomaly). The program replaced the Classics with a more practical education for future political students (Anomaly). It combines philosophical ethics, political systems, and economic modeling—three studies that are essential to a civil worker's toolkit. The degree creates a shift in the politician's mindset from reflection to leadership.

That approach is needed even more today. Modern leaders must be able to manage increasingly more complicated endeavors like global trade, AI, and fiscal policy—tasks that need more than historical insight. For example, a politician considering surveillance legislation must assess its ethical implications, financial impact, and political viability. The politician can't use historical knowledge in this scenario because legislation and scenarios like this have never occurred in history before. PPE prepares students to navigate these pressures. The program trains students to "provide future civil servants… with an opportunity to become generalists by exploring public policy through a blend of theory and application." (Anomaly).

While history grounds leaders, PPE equips them to act. It provides a toolkit designed for governance of the twenty-first century. It is tailored to let politicians handle current world problems like unemployment, climate policy, and fiscal crises, not just their historical parallels. In an age defined by complexity, speed, and interconnectedness, the most effective leaders are those who can think across disciplines. For that reason, PPE is not only a fitting education for modern political life—it is arguably essential.

Works Cited

Beckett, Andy. "PPE: The Oxford Degree That Runs Britain." *The Guardian*, 23 Feb. 2017, www.theguardian.com/education/2017/feb/23/ppe-oxford-university-degree-that-rules-britain.

Beschloss, Michael R. *The Conquerors*. 2002.

Burke, Olivia. digitalcommons.assumption.edu/cgi/viewcontent.cgi?article=1076&context=honorstheses.

College, Wheaton. "The Benefits of a PPE Certificate." *Wheaton College*, 2021, www.wheaton.edu/academics/academic-centers/wheaton-center-for-faith-politics-and-economics/resource-center/articles/2021/the-benefits-of-a-ppe-certificate/. Accessed 26 June 2025.

Dr. Jonathan Anomaly. "Why PPE?" *Political Science*, 29 Jan. 2016, polisci.duke.edu/news/why-ppe.

"Famous History Majors | History at Illinois." *History.illinois.edu*, history.illinois.edu/resources/careers/famous-history-majors.

Foster, David. *Writing and Learning in Cross-National Perspective*. Routledge, 2017.

L., Thomas. "Unveiling the Mysteries: Ancient China Unfolded." *Chronicles Uncovered: An Insight into Past Times*, 20 Sept. 2023, oralhistory.ws/resources/journey-through-ancient-china-a-historical-overview/.

Needleman, Rob. "The PPE Guide: Politics | UniAdmissions | World's First Oxbridge Prep School." *UniAdmissions | World's First Oxbridge Prep School*, 19 Apr. 2021, www.uniadmissions.co.uk/application-guides/the-ppe-guide-politics/. Accessed 12 June 2025.

OECD Forum For The Future, and Organisation For Economic Co-Operation And Development. *Governance in the 21st Century*. Organisation For Economic Co-operation and Development, 2001.

Putnam, Robert D., et al. *Making Democracy Work : Civic Traditions in Modern Italy*. Princeton Univ. Press, 1993.

Spring Hill College. *2022-2023 BULLETIN of INFORMATION*. pp. 150–50, www.shc.edu/wp-content/uploads/2022/09/2022-Bulletin-Traditional-Undergraduate-Programs.pdf. Accessed 26 June 2025.

"What Does PPE Stand for and Why Study a PPE Degree?" *Immerse Education*, 12 May 2025, www.immerse.education/university-preparation/choosing-a-degree/what-does-ppe-stand-for-and-why-study-a-ppe-degree/. Accessed 25 June 2025.

"Why Are History Majors Successful in Political and Public Policy Careers? | Department of History." *Osu.edu*, 2024, history.osu.edu/careers/politics.

Abstract

This paper studies short-term options with zero days to expiration (0DTE) in the QQQ and contrasts results with Brogaard, Han, and Won (2023) on the SPY. 0DTE options volume and composition have risen significantly over the past decade, from 14% of total QQQ linked options in the first quarter of 2011 to 43% in the third quarter of 2023. I find that 0DTE initially shows a positive correlation with volatility, but this association wanes once realized and implied volatility controls are included. I find that 0DTE serves as a proxy for broader volatility dynamics rather than driving the QQQ independently.

Options with zero days to expiration, abbreviated as 0DTE, exhibit elevated volatility due to gamma, which increases until expiration. This volatility is attractive to both retail "noise" traders as well as institutional participants in the form of potential returns; with the availability of 0DTE trading days extended to every business day, 0DTE volume has exponentially increased in the past decade (Table 1).

The introduction of short-term options by Nasdaq was as follows: weekly options expiring on Friday were introduced in late 2010, along with monthly options expiring on the third Friday of each month, and short dated quarterly options expiring on the last day of March, June, September, and December. Nasdaq announced the introduction of Monday and Wednesday weekly expirations in 2021 for QQQ options across its exchanges, including Nasdaq BX Options, NOM, ISE, GEMX, and MRX. CBOE and BOX Options Markets began listing Tuesday and Thursday weekly expirations for QQQ options in November 2022. Specifically, Tuesday-expiring weeklies started on November 14, 2022, and Thursday-expiring weeklies commenced on November 16, 2022.

The raw volume growth substantiates the boom of popularity in 0DTE trading, as shown in Graph 1:
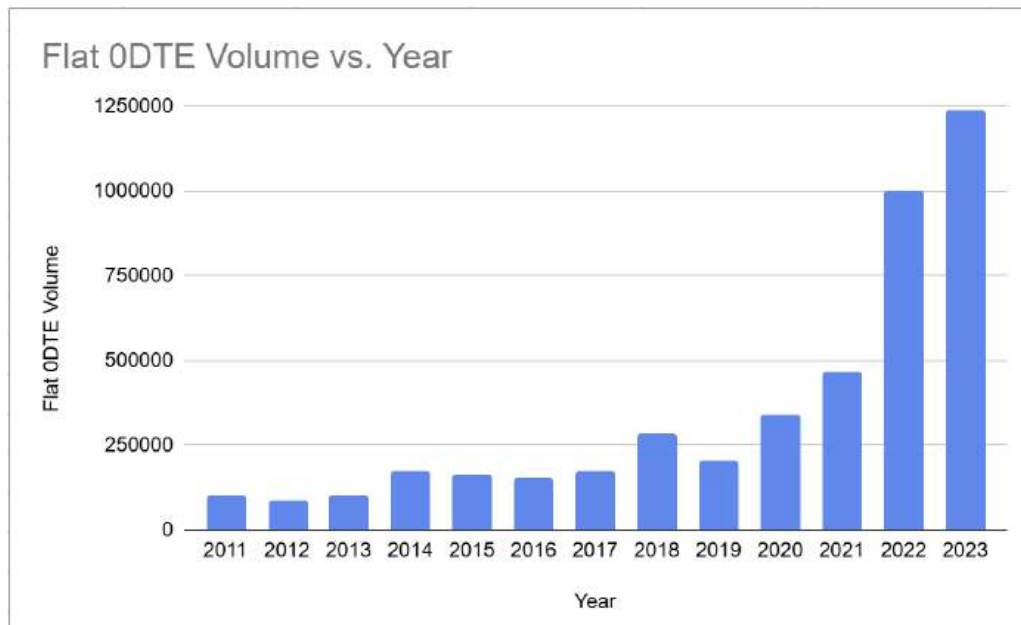
Total Volume of QQQ 0DTE Options, 2011-2023

QQQ's 0DTE volume expanded to over 1.2 million contracts by 2023, with particularly sharp acceleration beginning in 2020.

Notably, the S&P 500 was quicker to introduce 0DTE options than the Nasdaq. As a comparative analysis, I will reference the S&P 500 has a benchmark; more specifically, the research by Jonathan B Brogaard, Jaehee Han, and Peter Young Won in "Does 0DTE Options Trading Increase Volatility?" (2023). The differences between the two indexes, including but not exclusive to: composition, popularity, and institutional participation are vital to my hypothesis of how 0DTE's effect on volatility is more pronounced in the QQQ index.

The differences between indices start at their issuer; SPY tracks the S&P 500 while QQQ tracks the Nasdaq-100, with options traded on CBOE and Nasdaq exchanges respectively. While direct index options on SPX and NDX exist, their higher premium costs create barriers for retail participation. Therefore, SPY and QQQ serve as optimal subjects for 0DTE analysis due to their liquidity as the primary ETF representatives of their underlying indices and their greater accessibility to retail traders with limited capital.

QQQ's compositional differences with the SPY create the base of study. QQQ's technology concentration introduces sensitivity to earnings announcements and product

launches. Technology companies' global supply chains and operations also make QQQ more susceptible to overnight developments, creating information asymmetries exploitable with arbitrage. In contrast, the SPY, while still heavy in tech, is comprised with broader sector diversification that creates more gradual price movements and is less conductive to same day speculation as compared to the QQQ.

Retail investing popularity patterns further amplify QQQ's volatility; following Brogaard, Han, and Won's findings, I also contend that 0DTE options facilitate noise trading. The study of non-institutional (retail) traders' effects on volatility has been extensive (Shleifer and Vishny 1997; Kumar and Lee 2006; Lipson et al. 2023). De Long et al. (1990) argues that the inefficiency of predictions by institutional investors on retail traders can create a positive feedback loop of volatility. Poser (2023) presents how retail investors now constitute the majority of options trading activity, with their trading volume exceeding 50% concentrated in 0DTE options. Additionally, Bryzgalova et al. (2023) report that retail investors accounted for the majority of options trading activity, comprising over 60% of total volume during the period from November 2019 to June 2021. Covid stimuli and the rise of "meme" stocks and coins further exacerbate the volume of retail traders. These retail participants exhibit behavioral biases such as overconfidence in directional calls, preference for out-of-the-money options, and tendency toward momentum; traits grouped as speculative and uninformed trading (Liu et al. 2020). The psychological appeal of "betting on tech" during 0DTE sessions transforms hedging flows into speculative frenzies, particularly during earnings seasons or major tech announcements. I hypothesize that QQQ's technology weighting will serve as the primary explanatory variable for observed differences in 0DTE trading patterns.

Rather than examining absolute volume figures, which have grown alongside total options activity, I utilize 0DTE volume as a percentage of total daily options, or 0DTE%, to isolate the growing preference for same-day expiration strategies.
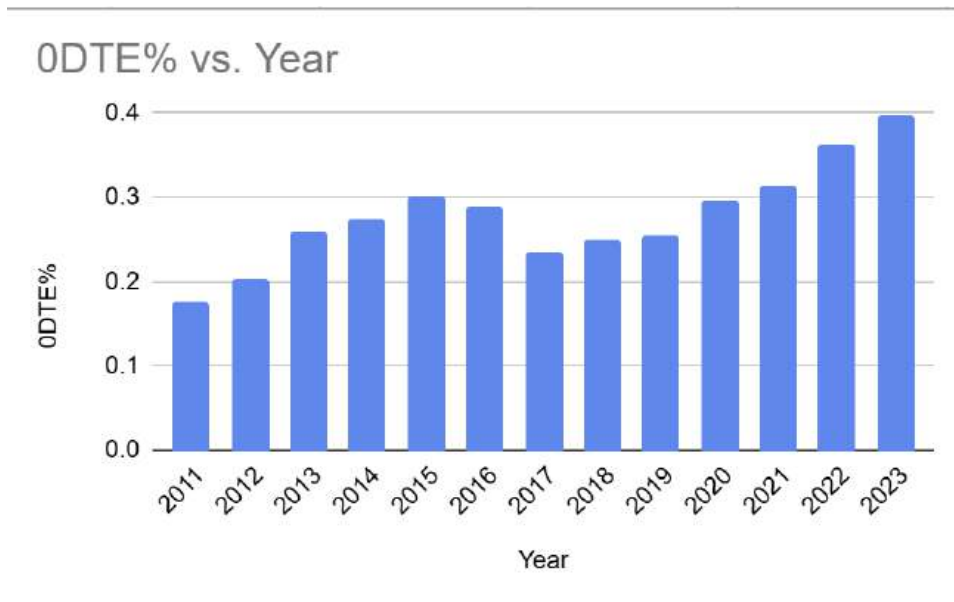
0DTE as Percent of Total QQQ Options by Year, 2011-2023

This percentage-based metric reveals QQQ's accelerating 0DTE adoption, rising from approximately 11% in January 2011 to nearly 40% by August 2023. This growth, concentrated in a technology-focused ETF, suggests that 0DTE activity's volatility effects are significant in QQQ.

Brogaard, Han, and Won (2023) serve as the primary foundation for this study. Their work will be cited and compared throughout, and readers are encouraged to review their original findings. Following the approach established by Brogaard et al. (2023), I use the Trade and Quote (TAQ) and OptionMetrics databases from WRDS, calculating both dependent and independent variables in the same manner. To calculate QQQ volatility, I pull the QQQ quote price from TAQ in the time period of January 2011 to August 2023. The calculation process samples bid-ask midpoints at five-minute intervals during standard trading hours (9:30 AM to 4:00 PM), yielding 78 observations per trading day. Daily volatility (σ), my independent variable, is derived by computing five-minute returns from these midpoints and calculating their standard deviation across each trading session. 0DTE% is my main dependent variable, and is expressed as the volume of QQQ 0DTE daily trades divided by the total volume of QQQ linked options traded that day.

I must note, that although the regressions and sources used for this study are identical in comparisons, the QQQ and SPY are fundamentally different indices. At the time of writing, the 10-year return of the QQQ is ~450% compared to the SPY at ~260%. However, the QQQ has a higher 1 month rolling volatility as well as a higher expense ratio, demonstrating QQQ's concentration in technology compared to SPY's wider diversification. Full summary statistics of all variables are included in table 7. I conjecture the differences between indices will result in results that are not necessarily aligned with Brogaard et al. (2023), despite similar sources and regressions in comparisons.

In the interval of January 2011 to August 2023, the number of yearly datapoints, or days where 0DTE traded, differs. The addition of Monday, Wednesday, Tuesday, and Thursday 0DTE affected the amount of datapoints, resulting in 2021, 2022, and 2023 having 125, 171, and 168 datapoints respectively (2023's datapoints end at 8/31/2023 as per OptionMetrics at the time of writing). This results in data weighted towards later dates, as there are more datapoints for later years with higher 0DTE volume than earlier years with lower 0DTE volume.

1. Data

Before exploring the differences between SPY and QQQ 0DTE trading, it's important to discuss some nuances with my dataset.

Following Brogaard et al. (2023), I utilize OptionMetrics as the source for the independent variable. However, there are irregularities in 0DTE% as compared to the SPY, a more liquid index. The years of 2011 – 2016 exhibit sustained elevated QQQ 0DTE% levels with no clear fundamental explanation, ranging from 10 - 15% higher than expected.

0DTE vs. Quarter
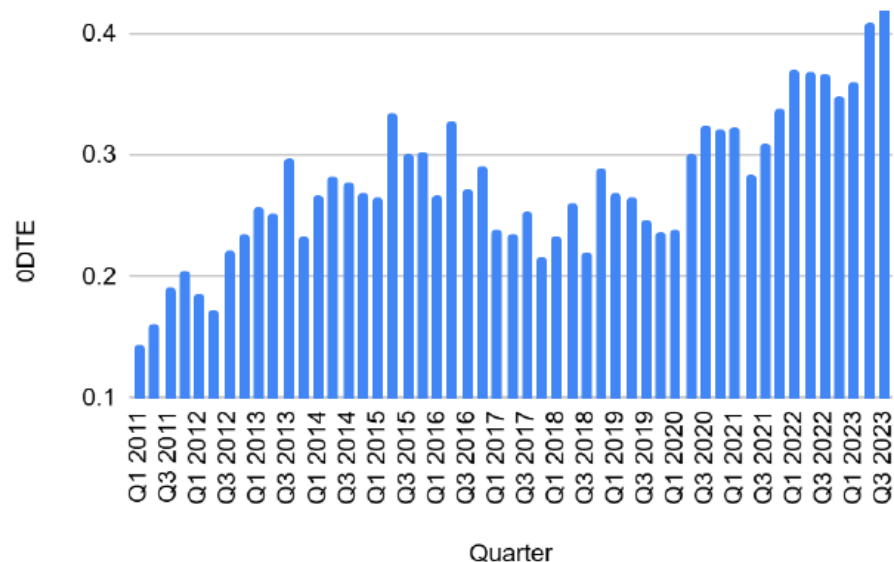
0DTE as Percent of Total QQQ Options by Quarter, 2011-2023
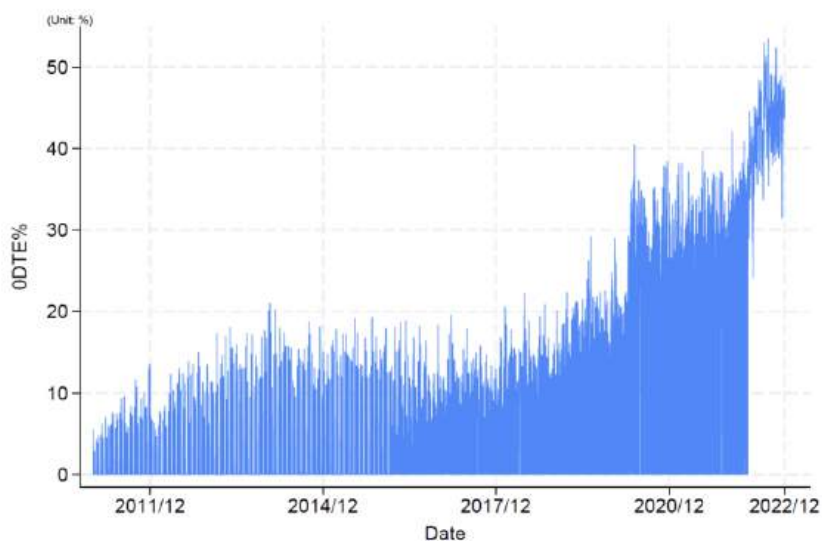
Figure 1: The trend of daily 0DTE%

This figure shows the daily trading volume of 0DTE as a percentage of the total trading volume for the S&P500 linked options (as denoted by 0DTE%) from January 2011 to December 2022.

"The trend of daily 0DTE%[SPY]", Brogaard et al. (2023)

I inquired to Optionmetrics through WRDS about the data irregularity, citing elevated 0DTE% over time despite patterns of similar outliers during the timeframe. OptionMetrics quotes this issue: "Unfortunately, due to the age of the data in question, we cannot confirm the possible inclusion of stale records or overstated volumes". OptionMetrics was able to replicate my findings but remained unable to fully validate the data quality for this earlier period. Their response stated: "Thank you for your feedback. We were able to make the necessary changes to our SQL query and have successfully replicated the user's data. Unfortunately, due to the age of the data in question, we cannot confirm the possible inclusion of stale records or overstated volumes. There is also no record of any nuances in IvyDB's treatment of QQQ option expiries in the 2011-2014 period".

Given these data quality concerns and the vendor's inability to confirm the accuracy of the early period data, I choose to begin the dataset in 2017. This point still captures several years before the 0DTE volume boom, while providing better data reliability. The sample extends to August 2023, the maximum available at the time of data capture, providing a sample size of 6 years, as opposed to 11 by Brogaard et al. (2023).

Additional data considerations include OptionMetrics' practice of taking snapshots at 3:59 PM (Bandi et al. 2023), potentially excluding options traded in the final minute before market close at 4:00 PM. This timing is relevant for 0DTE trading, as volatility typically increases as expiration approaches. However, I argue that the effect of this 1-minute absence may be limited due to executions of underlying securities requiring processing time, which likely occurs during this final minute.

The final dataset consists of 676 trading dates for both the OLS and IV regression, spanning January of 2017 to August 2023. When bid or ask quotes contain missing data for the independent variable, I apply forward fill to the data and maintain data continuity. While the quantity of trading dates does not directly affect the regression results, it does influence the overall sample size and statistical power of the analysis. I acknowledge that this dataset time difference will make regression comparison with Brogaard et al. (2023) more challenging.

I begin by regressing 0DTE% to QQQ market volatility; the regression analysis reveals that a one standard deviation increase in 0DTE volume percentage is associated with approximately a 7.921% increase in volatility in the QQQ. This finding presents a contrast to the

results reported by Brogaard, Han, and Won (2023), who found that a one standard deviation increases in 0DTE% corresponds to approximately a 9.03% increase in volatility for the SPY.

I employ an OLS and IV 2SLS regression for testing additional variables and robustness against endogeneity respectively. News and events, factors I weigh as more impactful to the QQQ due to composition, are represented in controls but are generalized and not considered individually, leading to omitted result bias that could pollute results. Reverse causality is another significant concern. The popularity of sports betting, retail trading, and cryptocurrency has introduced trends in volatility that could drive an increase in 0DTE volumes. To address these concerns and following Brogaard et al. (2023), I utilize a lagged 0DTE% from fifty business days ago, and as 0DTE expires on the day it is traded, it therefore cannot affect future volatility. The first stage of IV regression (Table 10), including only lagged 0DTE% and 0DTE%, shows the instrument is statistically significant, as evidenced by the Cragg-Donald F-statistics of 196.1.

2. Variables

In order to properly assess the effect to which 0DTE options trading has on index volatility, the implementation of control variables is justified and necessary.

Brogaard, Han, and Won (2023) serve as the primary foundation for this study. Their work will be cited and compared throughout, and readers are encouraged to review their original findings. Following the approach established by Brogaard et al. (2023), I use the Trade and Quote (TAQ) and OptionMetrics databases from WRDS, calculating both dependent and independent variables in the same manner. To calculate QQQ volatility, I pull the QQQ quote price from TAQ in the time period of January 2011 to August 2023. The calculation process samples bid-ask midpoints at five-minute intervals during standard trading hours (9:30 AM to 4:00 PM), yielding 78 observations per trading day. Daily volatility (σ), my independent variable, is derived by computing five-minute returns from these midpoints and calculating their standard deviation across each trading session. 0DTE% is the main dependent variable, and is expressed as the volume of QQQ 0DTE daily trades divided by the total volume of QQQ linked options traded that day.

Additional control variables following Brogaard et al. (2023) include: term spread, Default spread, Forex currency exchange rate, Consumer Price Index (CPI), Economic Policy Uncertainty Index (EPU), 22 business day rolling volatility (or 1 month volatility), and the daily

index return for the QQQ. These variables are added to account for established macroeconomic factors that influence volatility. Along with the foundational control variables, I introduce supplementary control variables that I hypothesize will capture differential sensitivities between the QQQ and SPY. The first of additional control variables includes the Geopolitical Risk Index (GPR), developed by Caldara, Dario, and Iacoviello (2021), which measures adverse geopolitical events and associated risks based on a tally of newspaper articles covering international tensions. This variable is particularly relevant given QQQ's technology composition, as semiconductor and hardware manufacturing require international supply chains that are extremely vulnerable to geopolitical disruptions compared to the broader economic sectors represented in SPY. Subsequently, I implement the Chinese Economic Policy Uncertainty Index (CHNEPU), developed by Baker, Bloom, Davis and Wang (2013), to capture production concerns from American technology's industrial competitor and trade partner, China (World bank).

Following international concern, I implement additional controls reflecting American sentiment. Firstly, the Monetary Policy Uncertainty Index (MPU), from the creators of the EPU, Baker-Bloom-Davis, captures the population's sentiment on the Fed. Nasdaq (2021) finds NDX has the highest (and) positive interest rate sensitivity with a correlation of 2.0, calculated as the price change per unit of interest rate change. The effects of Federal funds rate surprises on S&P 500 volatility and volatility risk premium compared to the S&P 500's correlation of only 0.7236. This sensitivity is consistent with "investors' interpretation of NDX as more growth and tech oriented." Secondly, I implement the VIX and VXN, a forward-looking index of sentiment over the next 30 days for the S&P500 and NDX respectively. These forward-looking measures help to mitigate the autocorrelation in my 22-day rolling volatility control, which looks to the past, while VXN provides index-specific expectations for QQQ's underlying Nasdaq-100 constituents and VIX captures broader market sentiment affecting SPY. Descriptive features of variables are portrayed in table 6.

Macroeconomic perception indexes such as the GPR, MPU, EPU, VIX, etc., often correlate and presents an issue of multicollinearity. I present a Pearson correlation matrix that presents the linear relationships between all variables in the analysis to assess the degree of correlation among these uncertainty measures.

```
===============================================================================================================================
CORRELATION MATRIX
-------------------------------------------------------------------------------------------------------------------------------
Variable           (1)      (2)     (3)      (4)      (5)    (6)    (7)     (8)     (9)    (10)    (11)    (12)    (13)   (14)
                dependent  0DTE%   Term   Default  Forex   CPI   22 R   Index   MPU    EPU    CHNEPU   GPR     VIX    VXN
-------------------------------------------------------------------------------------------------------------------------------
( 1) dependent     1.00
( 2) 0DTE%         0.11    1.00
( 3) Term Spread   0.03   -0.39    1.00
( 4) Default_Spread 0.26  -0.08   -0.01    1.00
( 5) Forex         0.00   -0.05    0.13   -0.14    1.00
( 6) CPI          -0.13    0.53   -0.07   -0.34    0.10   1.00
( 7) 22_rolling    0.44    0.20    0.06    0.53   -0.02  -0.12   1.00
( 8) Index_Return -0.10   -0.06   -0.05   -0.00    0.01   0.02  -0.08    1.00
( 9) MPU           0.26    0.22   -0.28    0.49   -0.13  -0.06   0.51   -0.03   1.00
(10) EPU           0.14    0.02   -0.03    0.63   -0.06  -0.03   0.41    0.01   0.32    1.00
(11) CHNEPU       -0.21   -0.02   -0.30   -0.13   -0.01  -0.08  -0.26    0.04  -0.25   -0.09    1.00
(12) GPR           0.18    0.25    0.16   -0.07   -0.10   0.12   0.32   -0.06   0.24   -0.06   -0.29    1.00
(13) VIX           0.44    0.18    0.15    0.59   -0.03   0.02   0.79   -0.16   0.52    0.52   -0.28    0.21    1.00
(14) VXN           0.47    0.26    0.11    0.50   -0.06   0.04   0.82   -0.18   0.52    0.45   -0.29    0.27    0.96   1.00
===============================================================================================================================
```

Correlation Matrix of Regression Control Variables

The strongest correlations appear between the uncertainty-related indexes, with the VIX and VXN showing a correlation coefficient of 0.96, indicating an extremely high positive relationship. Similarly, the EPU demonstrates strong correlations with MPU (0.32). Notably, CHNEPU trends inversely with U.S. policy and volatility measures, showing correlations of -0.25 with MPU, -0.09 with EPU, -0.28 with VIX, and -0.29 with VXN. In contrast, 0DTE% is more tied to macro conditions, showing a notable 0.53 with CPI and -0.39 with Term Spread. Most other variables remain loosely related, with Index Return largely independent.

The most substantial multicollinearity issue appears to be between VIX and VXN (0.96); this extremely high correlation is reasonable and expected. Given their conceptual similarity and near-perfect correlation, VIX and VXN will be used interchangeably in this analysis and will not appear together in the same regression model, thereby avoiding any multicollinearity issues while preserving the volatility information these indices provide. The addition of the VXN is to primarily serve as a better fit for a QQQ regression.

```
Summary Statistics
===============================================================================
Symbol                          N      Mean    Stdev    Min     Median    Max
===============================================================================
σt                              676    23.2    18.2     3.5     17.4     142.4

0DTE%                           676    32.9    7.9      9.8     33.4      51.4

Term Spread                     676    0.40    1.1      -1.73   0.58      2.28

Default Spread                  676    2.09    0.30     1.58    2.03      4.23

Forex                           676    0.30    0.11     0.27    0.28      0.99

CPI                             676    1.07    0.57     0.19    1.08      1.95

EPU                             676    1.37    0.84     0.17    1.17      6.34

22-day Rolling Volatility       676    1.37    0.68     0.365   1.23      5.62

Index Return                    676    0.054   1.43     -4.91   0.081     8.47

GPR                             676    1.10    0.41     0.58    1.05      3.19

MPU                             676    1.99    0.74     0.64    1.93      4.91

CHNEPU                          676    0.79    0.38     0.091   0.71      1.94

VIX                             676    19.96   6.86     9.14    18.8      66.0

VXN                             676    24.6    7.09     10.6    23.8      63.9
```

Summary statistics for all regression variables for the QQQ.

Several control variables in this study, namely CPI, MPU, GPR, and CHNEPU, are reported at monthly frequencies rather than daily. Each monthly observation for the aforementioned variables is repeated for all datapoints within the corresponding month. Additionally, to ensure a consistency of magnitudes across variables, several normalization adjustments have been applied to variables in the dataset. 0DTE% and 22-day rolling volatility measures have been multiplied by 100, consistent with Brogaard et al. (2023). Conversely, the uncertainty indexes, namely MPU, EPU, GPR, and CHNEPU, have been divided by 100. These normalizations do not affect underlying relationships between variables, and only move to improve readability of estimates.

The adjusted time frame of 2017-2023 selected introduces variability to the SPY. For instance, the CPI experienced extraordinary volatility during the COVID – 19 pandemic, with deflationary pressures followed by inflationary policies. A concentrated dataset covering this period will disproportionately weight these years of exceptional volatility. This is showcased by the QQQ having a mean CPI of 1.07 compared to the 0.62 in Brogaard et al. (2023). Similar

considerations apply to other macroeconomic variables; readers should attribute summary statistic differences not only to the QQQ but also the time frame studied.

When comparing statistics of the QQQ to that of the SPY, the independent variables' standard deviations, means, and maximum values are consistently higher for the QQQ. I conjecture that a significant portion of this elevated variation appears to stem from the midpoint price calculation methodology, which includes 1 datapoint out of 78 that calculates opening prices relative to the previous day's closing prices, heavily skewing the entire day's bid ask spreads, inflating volatility measurements. This effect is likely supported by the fact that many technology companies, which constitute a larger proportion of QQQ's holdings, typically release earnings announcements after markets close, creating greater overnight uncertainty and wider bid/ask spreads with the previously mentioned close-open calculation. Consequently, QQQ portrays more calculated volatility around shock events compared to the SPY, contributing to the observed differences in summary statistics.

3. QQQ 0DTE vs SPY 0DTE on Volatility
   A. OLS

The OLS regression equation employed both by this study and Brogaard et al. (2023) follows a log linear specification to examine the relationship between 0DTE volume and volatility.

The base model is expressed as follows:

$$Ln(\sigma) = \alpha + \beta(0DTE\%) + \delta'Xt + \varepsilon$$

where the $Ln(\sigma)$ represents the natural log of the standard deviation of midpoint bid-ask spreads for the QQQ during trading hours from 9:30 AM to 4:00 PM EST, calculated once daily. 0DTE, the main independent variable, measures the proportion of zero days to expiration options (options expiring the same day they are traded) relative to total daily option trading volume for the QQQ. Xt is a vector of independent controls, varying model to model. All standard errors are adjusted for heteroskedasticity and autocorrelation (HAC) by Newey-West (1994).

I employ eight model specifications to test robustness of the 0DTE-volatility relationship, while also controlling for omitted variable bias. The models progressively add control variables and are numbered from least to greatest in R-squared, starting with the baseline 0DTE specification and the standard controls mentioned alongside QQQ specific variables. Newey-West (1994) heteroskedasticity and autocorrelation - consistent (HAC) t-statistics (with optimal lag selection) are reported in parentheses below coefficients. The sample covers the period from January 2017 to August 2023.

```
FORMATTED RESULTS TABLE
```

Model #:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Comparison graphs:	(1)	(2)	(3)	(4)				
0DTE%	0.010 (2.011)	0.024 (5.41)	0.022 (4.88)	0.0097 (2.32)	0.0075 (1.89)	0.0085 (2.16)	0.0048 (1.27)	0.0059 (1.59)
Term Spread		0.07 (2.03)	0.030 (0.86)	0.019 (0.65)	-0.021 (-0.70)	-0.044 (-1.49)	-0.026 (-0.91)	-0.046 (-1.65)
Default Spread		0.53 (4.43)	0.44 (5.48)	0.15 (0.97)	-0.0025 (-0.015)	-0.13 (-0.98)	0.10 (0.66)	-0.032 (-0.26)
Forex		0.33 (1.29)	0.34 (1.52)	0.19 (0.76)	0.25 (1.15)	0.25 (1.25)	0.36 (1.71)	0.34 (1.82)
CPI		-0.23 (-3.06)	-0.26 (-3.57)	-0.14 (-2.14)	-0.21 (-3.34)	-0.26 (-4.34)	-0.21 (-3.20)	-0.25 (-4.18)
EPU		-0.004 (-0.11)		-0.046 (-1.27)	-0.077 (-2.07)		-0.076 (-2.06)	
CHNEPU			-0.34 (-3.80)			-0.23 (-3.03)		-0.22 (-2.82)
22 Rolling				0.39 (5.99)	0.16 (2.12)	0.15 (2.04)	0.076 (0.93)	0.075 (0.94)
Index Return				-0.028 (-1.74)	-0.0099 (-0.67)	-0.011 (-0.75)	-0.0056 (-0.37)	-0.0071 (-0.46)
MPU					-0.015 (-0.34)	-0.044 (-1.02)	-0.024 (-0.58)	-0.049 (-1.22)
GPR					0.099 (1.51)	0.076 (1.21)	0.103 (1.67)	0.082 (1.36)
VIX					0.035 (5.30)	0.033 (5.24)		
VXN							0.041 (6.10)	0.038 (5.87)
Observations	676	676	676	676	676	676	676	676
Adj R-squared	0.011	0.111	0.141	0.206	0.236	0.244	0.249	0.255

Multi-Model OLS Analysis of 0DTE Options Volume and QQQ Volatility with Controls

Table 8 as shown above presents results of all 8 regressions. The univariate (col.1, model 1) shows a positive 0DTE% to volatility link (β=0.010; t=2.01). Adding the baseline macro controls but still omitting 22-day rolling volatility and index return in model 2 amplifies the 0DTE% coefficient to 0.024 (t =5.41). Once CHNEPU/EPU are introduced in model 3, the effect moderates (0.022; t =4.88). Notably, CHNEPU has a more pronounced effect on QQQ volatility compared to the stablished EPU. Bringing in 22-day rolling volatility and index return

in model 4 lowers 0DTE%'s coefficient further to 0.0097 (t =2.39). With broader uncertainty and volatility proxies added in models 5-8, the 0DTE% coefficient falls into the ~0.0048–0.0085 range and loses conventional significance in the VXN-specs (cols. 7–8 t = 1.27–1.59). Notably, in the equivalent baseline in model 4, QQQ's 0DTE% effect (0.0097) exceeds the SPY estimate of 0.007. As expected, volatility indexes are powerful; VIX is positive and highly significant when used (cols. 5–6; t = 5.2–5.3), and VXN fits the QQQ regressions even better (cols. 7–8; t = 5.9–6.1), coinciding with the highest adj. R squared. CPI is consistently negative and significant across most columns; 22-day rolling vol is positive and significant where included; other controls (term spread, forex, EPU, GPR, index return) are generally weaker or only intermittently significant. The attenuation of the 0DTE% as 22-day rolling volatility and VIX/VXN controls are included suggests that 0DTE activity is, at least in part, serving as a proxy for a broader volatility regime. Once those regime defining measures are accounted for, the explanatory power of 0DTE% diminishes. I speculate that 0DTE's predictive power is tied to underlying market volatility conditions rather than being an independent driver.

Table 9 reports the OLS VIF values across all eight models and shows that multicollinearity is generally not a concern. Most regressors remain comfortably below the threshold of 5, with only VIX (4.0–4.2) and VXN (4.1) approaching it. 0DTE% itself stays low (1.7–2.1), suggesting its estimated effects are not distorted by collinearity.

B. IV (2SLS) Regression

The methodological considerations by Brogaard et al. (2023) establish that their 0DTE% variable satisfies stationarity requirements through unit root testing. Calculating both Augmented Dickey-Fuller and Phillips-Perron tests, they present statistical rejection of non-stationarity at the 1% significance level, with test statistics of -34.99 and -44.63 respectively. The results provide solid precedent for my own volatility measure validation, as the statistical properties of variables must be confirmed before proceeding with the regression. Of particular relevance to my study is their treatment of autocorrelation concerns in volatility measures. Brogaard et al. (2023) portray that volatility exhibits no statistically significant autocorrelation at the 50-business-day lag using Bartlett's Moving Average confidence intervals. They show the autocorrelation estimate falls within acceptable bounds, validating the exclusion restriction required. They further test this with

alternative volatility measures (50-day lagged VIX and realized volatility) finding insignificant coefficients of -0.040 and -0.501 respectively.

These findings directly support my methodological approach. Since I mirror the same equation structure as Brogaard et al. (2023) and apply it to the QQQ index, their validation that lagged volatility measures do not create endogeneity problems provides justification for the basis of this regression. The stationarity and autocorrelation properties they establish should similarly hold for my QQQ-based analysis.

The IV regression equation employed both by this study and Brogaard et al. (2023) utilizes a lagged 0DTE of 50 days to examine the relationship between 0DTE volume and volatility.

The regression equation for the first stage regression is as follows:

$$0DTE\% = \alpha + \beta 0DTE\%_{t-50} + \delta'X_t + \gamma_t + u_t$$

where 0DTE%t-50 is 0DTE% from 50 business days ago and Xt a vector for control variables varying from model to model. All standard errors adjusted for HAC.

The regression equation for the second stage regression is as follows:

$$L(\sigma_t) = \alpha + \beta\ 0DTE^*\%_t + \delta'X_t + \gamma_t + \varepsilon_t$$

where the Ln(σ) represents the natural log of the standard deviation of midpoint bid-ask spreads for the QQQ during trading hours from 9:30 AM to 4:00 PM EST, calculated once daily. 0DTE*%t is the transformed 0DTE% from the first regression, and Xt a vector for control variables varying from model to model. All standard errors adjusted for HAC.

The 2SLS Multi-Model IV (Table 10) is presented below. Models 1-5 represent the first-stage regressions, while models 6-10 are the second-stage regressions. The Cragg-Donald F statistics are all significantly above the conventional threshold of 10, confirming the relevance and strength of the instruments. However, the instrument strength is weaker than that reported in Brogaard et al. (2023), where the average F statistic across three models was ~240. This

difference in instrument strength suggests that while the instruments remain statistically valid, the precision of the IV estimates lead to be reduced relative to the benchmark study.

		First Stage					Second Stage			
Model#	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0DTE%_{t-50}	0.505***	0.113**	0.205***	0.070	0.134***					
	(9.98)	(2.21)	(4.21)	(1.47)	(3.16)					
0DTE%*						0.006	-0.015	0.048**	-0.100	0.023
						(0.52)	(-0.37)	(2.01)	(-1.02)	(0.66)
Term Spread_t		-2.389***	-2.109***	-2.907***	-2.475***		-0.046	0.127*	-0.346	0.077
		(-7.99)	(-5.89)	(-9.83)	(-7.69)		(-0.41)	(1.92)	(-1.13)	(0.76)
Default Spread_t		-0.382	3.213**	0.732	3.157**		0.136	0.438***	0.169	0.375**
		(-0.23)	(2.14)	(0.46)	(2.04)		(0.81)	(2.79)	(0.65)	(2.01)
σ^forex_t		-3.257***	-1.966	-0.994	0.124		0.097	0.402	0.234	0.454*
		(-3.05)	(-1.38)	(-0.81)	(0.09)		(0.34)	(1.55)	(1.04)	(1.84)
σ^cpi_t		6.887***	6.535***	6.423***	6.478***		0.044	-0.417**	0.506	-0.262
		(11.41)	(8.94)	(11.08)	(10.04)		(0.14)	(-2.20)	(0.76)	(-1.01)
22_rolling		3.493***		1.257		0.491***			0.211	
		(4.86)		(1.22)		(2.70)			(1.02)	
EPU_t		-0.877**	-0.530	-0.936**	-0.439		-0.07	0.012	-0.181	0.003
		(-2.19)	(-1.30)	(-2.51)	(-1.15)		(-1.28)	(0.25)	(-1.60)	(0.07)
Index Return_t		-0.362**	-0.483***	-0.218	-0.411***		-0.037	-0.023	0.027	-0.030
		(-2.35)	(-3.15)	(-1.33)	(-2.77)		(-1.55)	(-1.18)	(-0.85)	(-1.42)
MPU_t				-0.748	0.195				-0.100	0.121**
				(-1.52)	(0.37)				(-0.93)	(2.44)
GPR_t				3.221***	4.231***				0.460	0.169
				(4.10)	(4.82)				(1.34)	(0.93)
VXN				0.222**					0.066**	
				(2.39)					(2.46)	
Observations	676	676	676	676	676	676	676	676	676	676
Adjusted R-squared						0.009	0.164	0.068	-0.454	0.141
Cragg-Donald F-statistic	196.1	85.0	77.5	71.9	71.5					

Multi-Model IV 2SLS Regression on 0DTE and Volatility with Controls

In the same setup specifications with Brogaard et al. (2023), both studies show strong instrument relevance, although SPY's is stronger in magnitude. My first stage loading on LAG50 $0DTE\%$ is 0.113 compared to Brogaard et al (2023)'s 0.237. The results diverge in the second stage. Brogaard et al (2023)'s SPY regressions produce a positive and significant 0DTE coefficient at 0.019 with a t statistic of 2.91, while my QQQ analogue yields a similar in magnitude but a negative and insignificant 0DTE* (instrumented 0DTE) estimate of -0.015 with a t statistic of -0.37. I interpret this IV result not necessarily as evidence against a correlation between 0DTE and volatility, but rather that in the QQQ, 0DTE* is highly collinear with volatility variables in the model, particularly realized and implied volatility, so its incremental contribution is much harder to isolate once volatility controls are present.

To expand on this, I built models 8 through 10, corresponding to columns 8 through 10, specifically to test whether 0DTE* provides anything beyond serving as a proxy for volatility. In Model 8, I remove the 22-day rolling realized volatility control while keeping the rest of the macro variables intact. In Model 9, I include the full control set from the OLS regression, opting

to include VXN instead of the VIX as per the stronger fit demonstrated in the OLS regression. Finally, in model 10, I drop both volatility measures, 22-day rolling volatility and VXN, so that any residual 0DTE* effect must emerge without those channels.

When I omit 22-day rolling volatility in model 8, 0DTE* turns positive and reaches significance with an estimate of 0.048 and a t statistic of 2.01, which is what would be anticipated if realized volatility had been capturing the same signal. At the same time, the term and default spreads lean in the positive direction (0.127, t = 1.92; 0.438, t = 2.79), while CPI comes in negative and significant (−0.417, t = −2.20). In the control heavy model 9, the 0DTE* coefficient flips negative and loses all precision at an estimate of -0.100, t = -1.02. The model suffers from multicollinearity across volatility proxies and sentiment indices, evident in EPU (−0.181, t = −1.60) and noisy GPR (0.460, t = 1.34). In model 10, with both 22-day rolling volatility and VXN removed, 0DTE* again becomes small and insignificant with an estimate of 0.023 and a t statistic of 0.66. What is most interesting here is that the other risk proxies step forward; default spread is materially positive (0.375, t = 2.01), FX volatility edges up (0.454, t = 1.84), and MPU becomes significant (0.121, t = 2.44).

These results support my previous prediction that 0DTE activity primarily functions as a proxy for broader volatility dynamics rather than exerting an independent causal effect on underlying market volatility. The only model where 0DTE* reaches significance is the one that excludes realized volatility; once either realized volatility or implied volatility are included, 0DTE*'s estimate becomes unstable and insignificant while volatility controls carry the explanatory weight. The disappearance of 0DTE significance upon removing volatility-related controls indicates that the relationship functions through volatility rather than through a path unique to same-day options. To explore this further, I present robustness tests to prove whether any directional causality exists beyond these proxy relationships.

C. Robustness

The orthogonalization analysis tests whether 0DTE retains predictive power once its variation explained by volatility measures is stripped out. I estimate models in which 0DTE is orthogonalized against both realized 22-day rolling volatility and implied volatility (VXN), as well as against each measure separately. The orthogonalization is presented below.

```
================================================================================
SUMMARY TABLE: Comparison of All Models (Newey-West HAC Standard Errors)
================================================================================
Model 1: Orthogonalization with both 22-Rolling and VXN
Model 2: Orthogonalization with 22-Rolling
Model 3: Orthogonalization with VXN
================================================================================
Variable            Model 1                Model 2                Model 3
--------------------------------------------------------------------------------
0DTE%               -0.0004 (-0.08)         0.0032 (0.60)         -0.0008 (-0.15)
CPI                 -0.0399 (-0.49)        -0.0609 (-0.74)        -0.0379 (-0.46)
Default_Spread       0.6048*** (4.72)       0.6021*** (4.72)       0.6047*** (4.72)
EPU                 -0.0186 (-0.44)        -0.0158 (-0.38)        -0.0189 (-0.45)
Index_Return        -0.0463*** (-2.62)     -0.0465*** (-2.64)     -0.0462*** (-2.63)
Term_Spread          0.0117 (0.32)          0.0184 (0.51)          0.0110 (0.30)
const                1.7043*** (6.04)       1.7260*** (6.11)       1.7031*** (6.03)
--------------------------------------------------------------------------------
Observations        676                    676                    676
Adjusted R-squared  0.0733                 0.0741                 0.0733
Durbin-Watson       1.4177                 1.4187                 1.4178
```

Orthogonalization of 0DTE with 22-Day Rolling Volatility and VXN

The orthogonalization results give a clearer picture of how lagged 0DTE interacts with different volatility measures. When both VXN and 22-day rolling volatility are included, the fit is modest, with an R squared of 0.0815 and an adjusted R squared of 0.0733, though the F statistic of 12.11 confirms strong overall significance. Using only the 22-day rolling volatility measure produces nearly the same fit, with an R squared of 0.0823 and adjusted R squared of 0.0741, and the small positive coefficient of 0.0032 suggests realized volatility overlaps with, but does not erase, the role of 0DTE. The picture changes slightly with VXN alone, where the fit is weaker at an R squared of 0.0816 and the coefficient turns negative at minus 0.0008, pointing to implied volatility as a closer driver of lagged 0DTE. I employ a Vector Autoregression framework with impulse response functions that show how shocks propagate through the system, supplemented by forecast error variance decomposition analysis (FEVD) in the appendix.

The most striking differences across the three VAR-IRF specifications (Graphs 13-15) appear in the 0DTE → dependent impulse response panels. In the raw 0DTE model, the dependent variable shows virtually no response to 0DTE shocks, remaining flat around zero throughout the 10-period horizon with tight confidence bands. The lag 50 0DTE specification exhibits a more pronounced pattern, with the dependent variable displaying an initial negative response that reaches approximately -0.05 before gradually returning toward zero, though the confidence bands suggest this may not be statistically significant. The orthogonalized model

shows the most dramatic response, with an immediate sharp drop to around -0.10 in the first period, followed by continued negative values that persist through the impulse horizon. The confidence bands in this final specification appear wider, indicating greater uncertainty, but the magnitude of the response is notably larger than in either of the previous two models.

The impulse response patterns become even clearer when examining how much each variable contributes to forecast error variance over time through the FEVD analysis (Table 16). In the raw 0DTE specification, the dependent variable (QQQ volatility) shows remarkable independence - 0DTE explains virtually none of its forecast error variance across all periods (maxing out at 0.48% in period 1), while the dependent variable itself accounts for 99.62% initially before declining to 96.78% by period 1 as VXN's influence grows to 2.74%. The lag 0DTE results tell a similar story, with 0DTE's contribution to dependent variable variance remaining minimal at just 1.27% in period 1 and declining to 0.09% by period 2, while VXN maintains a modest but consistent role around 2-3%. The orthogonalized specification reveals the most dramatic shift in explanatory power - here, 0DTE's contribution jumps significantly to 2.11% in period 1 and actually increases to 2.00% by period 2, representing roughly a four-fold increase compared to the other specifications. Notably, across all three models, VXN consistently dominates its own forecast error variance (maintaining 94-97% throughout), while 0DTE shows increasing ability to explain VXN variance in the orthogonalized case (2.08% in period 1, rising to 2.51% by period 2).

```
=== Combined Robustness Table (VIF + VAR Correlations) ===

          Case  Variable       VIF  Corr_0DTE  Corr_dependent   Corr_VXN
          Raw      const 33.202064        NaN             NaN        NaN
          Raw       0DTE  1.075085   1.000000        0.062132  -0.011303
          Raw dependent  1.281616   0.062132        1.000000   0.151019
          Raw        VXN  1.366624  -0.011303        0.151019   1.000000
          Lag      const 36.795438        NaN             NaN        NaN
          Lag       0DTE  1.073701   1.000000       -0.043067  -0.019320
          Lag dependent  1.293532  -0.043067        1.000000   0.152482
          Lag        VXN  1.374408  -0.019320        0.152482   1.000000
 Orthogonalized     const 21.699427        NaN             NaN        NaN
 Orthogonalized      0DTE  1.008381   1.000000       -0.064580  -0.163121
 Orthogonalized dependent  1.292064  -0.064580        1.000000   0.152138
 Orthogonalized       VXN  1.283682  -0.163121        0.152138   1.000000
```

Robustness Appendix of VIF and VAR correlations

The robustness diagnostics reveal important insights about multicollinearity and residual correlations across the three VAR specifications. The VIF values remain well below conventional thresholds of concern across all models, with the highest VIF of 36.8 appearing for the constant term in the lag specification, while variable VIFs stay comfortably in the 1.0-1.4 range, indicating minimal multicollinearity issues. The correlation patterns between residuals show meaningful variation across specifications - in the raw 0DTE model, the correlation between 0DTE and the dependent variable is positive at 0.062, while the 0DTE-VXN correlation is slightly negative at -0.011. The lag specification maintains similar patterns with a negative 0DTE-dependent correlation of -0.043 and a slightly stronger negative 0DTE-VXN correlation of -0.019. Most notably, the orthogonalized specification shows the strongest negative correlations, with 0DTE-dependent at -0.065 and 0DTE-VXN reaching -0.163, suggesting that the orthogonalization process successfully isolated distinct variance components. The dependent-VXN correlations remain remarkably stable around 0.15 across all three models, indicating a consistent underlying relationship between QQQ volatility and VXN that persists regardless of how 0DTE is treated. These diagnostic results confirm that the observed differences in impulse responses and variance decompositions reflect genuine structural relationships rather than statistical artifacts from multicollinearity or model misspecification.

4. Conclusion

I build on Brogaard, Han, and Won (2023) and their analysis on 0DTE options volume on the SPY by extending controls and applying methodologies to the QQQ. Brogaard et al. (2023) document a strong and positive relationship between 0DTE options volume and SPY, and I assess whether the composition of the QQQ and investor behavior can influence volatility dynamics differing from the SPY through 0DTE. Applying identical methodology to the QQQ of using a lag 50 0DTE, I calculate $0DTE\%$ from 2017 to 2023 and measure daily volatility from intraday midpoint bid/ask spreads. OLS regressions with variable controls indicate a positive link between $0DTE\%$ and QQQ volatility, which slightly exceeds the SPY baseline in comparable specifications, and is consistent with the technology focused composition of the QQQ. However, the IV regressions with variable controls shows a weaker instrument strength and an insignificant 0DTE* effect.

Orthogonalization of 0DTE confirms that it's predictive power largely overlaps with broader volatility measures, while VAR IRF and FEVD analyses showcase that 0DTE options explain only a minor portion of forecast error variance when against VXN and 22-day rolling volatility. I conjecture that the divergence between the SPY and QQQ can be attributed to structural differences with QQQ's greater technology weighting amplifying sensitivity to earnings announcements, international tensions, product launches, and retail sentiment. Moreover, macro sentiment controls interact with volatility, further attenuating 0DTE's independent effect. Collectively, these findings suggest that 0DTE trading contributes to elevated intraday volatility, but its role is context dependent, functioning more as a volatility proxy in the QQQ instead of a distinct driver, in contrast to SPY's independent 0DTE.

5. Figures

Graph 1: Total Volume of QQQ 0DTE Options, 2011-2023
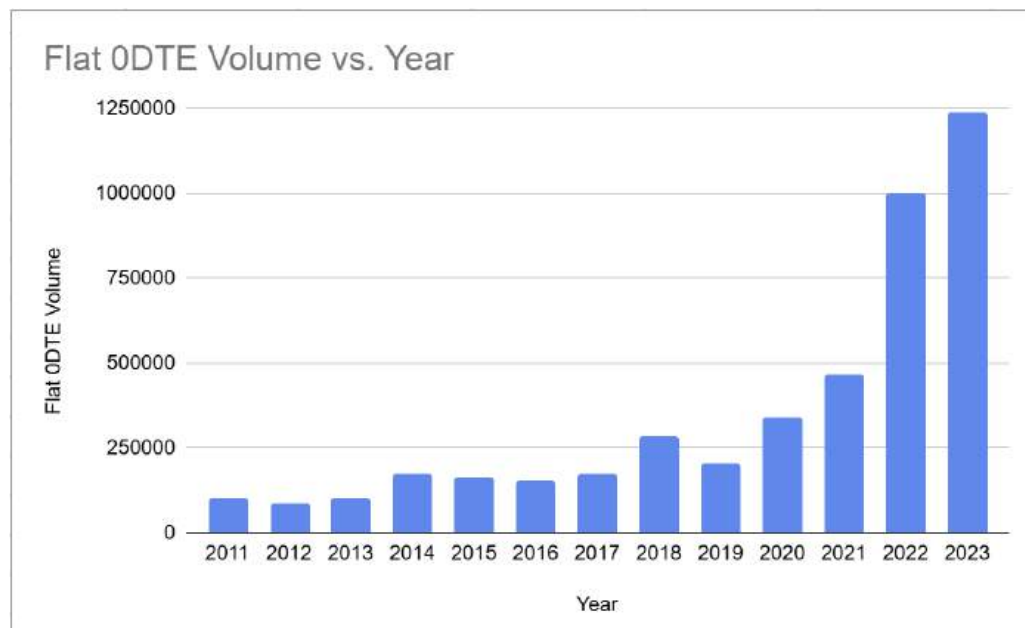
Graph 2: 0DTE as Percent of Total QQQ Options by Year, 2011-2023

0DTE% vs. Year
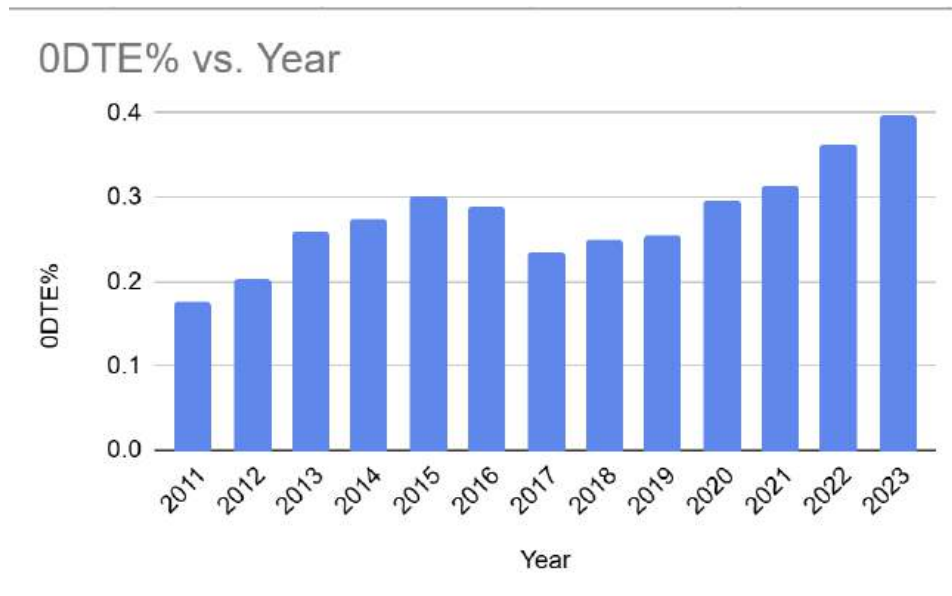
Table 3: 0DTE as Percent of Total QQQ Options by Quarter, 2011-2023

0DTE vs. Quarter
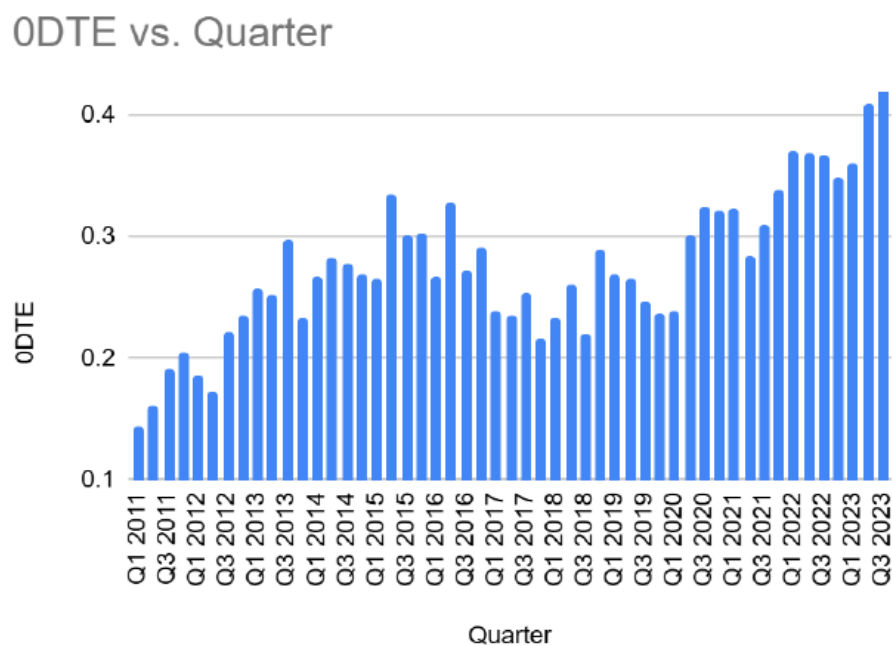
Table 4: "The trend of daily 0DTE%[SPY]",

"This figure shows the daily trading volume of 0DTE as a percentage of the total trading volume for the S&P500 linked options (as denoted by 0DTE%) from January 2011 to December 2022" Brogaard et al. (2023)
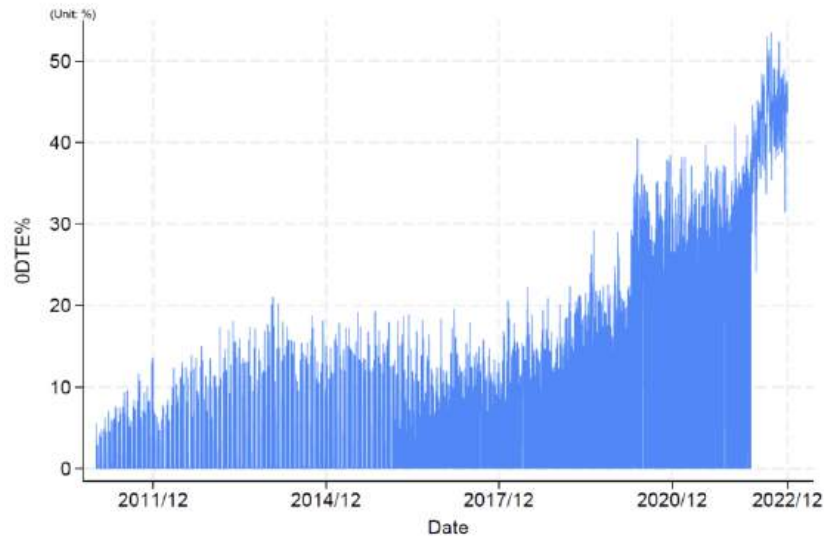
Figure 1: The trend of daily 0DTE%
This figure shows the daily trading volume of 0DTE as a percentage of the total trading volume for the S&P500 linked options (as denoted by 0DTE%) from January 2011 to December 2022.

Table 5: Correlation Matrix of Regression Control Variables

Table 6: Description of Regression Control Variables

```
================================================================
Symbol                     Definition
================================================================
σt                         The daily standard deviation of the Invesco QQQ Trust (QQQ) midpoint quote return every 5
                           minutes between 9:30 AM and 4:00 PM Eastern Time. The QQQ tracks the Nasdaq-100.

0DTE%                      The daily QQQ zero days to expiration options (0DTE) trading volume divided by daily total
                           QQQ related options volume.

Term Spread                The daily difference between long term yield on government bonds and the Treasury bill rate.
                           Specifically, Term Spread is calculated as the 10 year Treasury Maturity rate minus the
                           3 month Treasury bill secondary market rate.

Default Spread             The daily difference between the yield on BAA-rated corporate bonds and the yield on long
                           term US government bonds.

Forex                      The standard deviation of the daily return of the foreign exchange rates over the previous
                           22 day rolling periods.

CPI                        The standard deviation of the previous 12 month returns based on the year-over-year returns
                           of the Consumer Price Index.
                           Calculated as percent change: (CPI at current month - CPI at the  same month last year) /
                           CPI at same month last year .

EPU                        The daily Economic Policy Uncertainty (EPU) index in Baker et al. (2016).

22-day Rolling Volatility  The standard deviation of the daily returns of the QQQ over the previous 22-day rolling periods.

Index Return               The daily return of the Invesco QQQ Trust (QQQ). Expressed in a percent.

GPR                        The daily Geopolitical Risk Index (GPR) index by Caldara, Dario, and Iacoviello (2021).

MPU                        The daily Monetary Policy Uncertainty index (MPU) developed by Baker, Bloom, and Davis.

CHNEPU                     The daily Chinese Economic Policy Uncertainty index (CHNEPU) by Baker, Bloom, Davis and Wang.

VIX                        The daily CBOE volatility index. Calculated for the S&P 500.

VXN                        The daily CBOE volatility index. Calculated for the Nasdaq-100.
```

Table 7: Summary Statistics of Regression Control Variables

Summary Statistics

Symbol	N	Mean	Stdev	Min	Median	Max
σt	676	23.2	18.2	3.5	17.4	142.4
0DTE%	676	32.9	7.9	9.8	33.4	51.4
Term Spread	676	0.40	1.1	-1.73	0.58	2.28
Default Spread	676	2.09	0.30	1.58	2.03	4.23
Forex	676	0.30	0.11	0.27	0.28	0.99
CPI	676	1.07	0.57	0.19	1.08	1.95
EPU	676	1.37	0.84	0.17	1.17	6.34
22-day Rolling Volatility	676	1.37	0.68	0.365	1.23	5.62
Index Return	676	0.054	1.43	-4.91	0.081	8.47
GPR	676	1.10	0.41	0.58	1.05	3.19
MPU	676	1.99	0.74	0.64	1.93	4.91
CHNEPU	676	0.79	0.38	0.091	0.71	1.94
VIX	676	19.96	6.86	9.14	18.8	66.0
VXN	676	24.6	7.09	10.6	23.8	63.9

Table 8: Multi-Model OLS Analysis of 0DTE Options Volume and QQQ Volatility with Controls

```
===============================================================================================================
FORMATTED RESULTS TABLE
===============================================================================================================
Model #:            (1)       (2)       (3)       (4)       (5)       (6)       (7)       (8)
Comparison graphs:  (1)       (2)                 (4)
---------------------------------------------------------------------------------------------------------------
0DTE%             0.010     0.024     0.022     0.0097    0.0075    0.0085    0.0048    0.0059
                 (2.011)   ( 5.41)   ( 4.88)   ( 2.32)   ( 1.89)   ( 2.16)   ( 1.27)   ( 1.59)
---------------------------------------------------------------------------------------------------------------
Term Spread                 0.07      0.030     0.019    -0.021    -0.044    -0.026    -0.046
                           ( 2.03)   ( 0.86)   ( 0.65)   (-0.70)   (-1.49)   (-0.91)   (-1.65)
---------------------------------------------------------------------------------------------------------------
Default Spread              0.53      0.44      0.15     -0.0025   -0.13      0.10     -0.032
                           ( 4.43)   ( 5.48)   ( 0.97)  (-0.015)  (-0.98)   ( 0.66)   (-0.26)
---------------------------------------------------------------------------------------------------------------
Forex                       0.33      0.34      0.19      0.25      0.25      0.36      0.34
                           ( 1.29)   ( 1.52)   ( 0.76)   ( 1.15)   ( 1.25)   ( 1.71)   ( 1.82)
---------------------------------------------------------------------------------------------------------------
CPI                        -0.23     -0.26     -0.14     -0.21     -0.26     -0.21     -0.25
                           (-3.06)   (-3.57)   (-2.14)   (-3.34)   (-4.34)   (-3.20)   (-4.18)
---------------------------------------------------------------------------------------------------------------
EPU                        -0.004              -0.046    -0.077              -0.076
                           (-0.11)             (-1.27)   (-2.07)             (-2.06)
---------------------------------------------------------------------------------------------------------------
CHNEPU                                -0.34                        -0.23               -0.22
                                     (-3.80)                       (-3.03)             (-2.82)
---------------------------------------------------------------------------------------------------------------
22 Rolling                                      0.39      0.16      0.15      0.076     0.075
                                               ( 5.99)   ( 2.12)   ( 2.04)   ( 0.93)   ( 0.94)
---------------------------------------------------------------------------------------------------------------
Index Return                                   -0.028    -0.0099   -0.011    -0.0056   -0.0071
                                               (-1.74)   (-0.67)   (-0.75)   (-0.37)   (-0.46)
---------------------------------------------------------------------------------------------------------------
MPU                                                      -0.015    -0.044    -0.024    -0.049
                                                         (-0.34)   (-1.02)   (-0.58)   (-1.22)
---------------------------------------------------------------------------------------------------------------
GPR                                                       0.099     0.076     0.103     0.082
                                                         ( 1.51)   ( 1.21)   ( 1.67)   ( 1.36)
---------------------------------------------------------------------------------------------------------------
VIX                                                       0.035     0.033
                                                         ( 5.30)   ( 5.24)
---------------------------------------------------------------------------------------------------------------
VXN                                                                           0.041     0.038
                                                                             ( 6.10)   ( 5.87)
---------------------------------------------------------------------------------------------------------------
Observations      676       676       676       676       676       676       676       676
Adj R-squared     0.011     0.111     0.141     0.206     0.236     0.244     0.249     0.255
```

Table 9: VIF Analysis of Multi-Model OLS

Table 10: Multi-Model IV 2SLS Regression on 0DTE and Volatility with Controls

| | First Stage | | | | | Second Stage | | | | |
Model#	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0DTE%_{t-50}	0.505***	0.113**	0.205***	0.070	0.134***					
	(9.98)	(2.21)	(4.21)	(1.47)	(3.16)					
0DTE%*						0.006	-0.015	0.048**	-0.100	0.023
						(0.52)	(-0.37)	(2.01)	(-1.02)	(0.66)
Term Spread_t		-2.389***	-2.109***	-2.907***	-2.475***		-0.046	0.127*	-0.346	0.077
		(-7.99)	(-5.89)	(-9.83)	(-7.69)		(-0.41)	(1.92)	(-1.13)	(0.76)
Default Spread_t		-0.382	3.213**	0.732	3.157**		0.136	0.438***	0.169	0.375**
		(-0.23)	(2.14)	(0.46)	(2.04)		(0.81)	(2.79)	(0.65)	(2.01)
σ^forex_t		-3.257***	-1.966	-0.994	0.124		0.097	0.402	0.234	0.454*
		(-3.05)	(-1.38)	(-0.81)	(0.09)		(0.34)	(1.55)	(1.04)	(1.84)
σ^cpi_t		6.887***	6.535***	6.423***	6.478***		0.044	-0.417**	0.506	-0.262
		(11.41)	(8.94)	(11.08)	(10.04)		(0.14)	(-2.20)	(0.76)	(-1.01)
22_rolling		3.493***		1.257			0.491***		0.211	
		(4.86)		(1.22)			(2.70)		(1.02)	
EPU_t		-0.877**	-0.530	-0.936**	-0.439		-0.07	0.012	-0.181	0.003
		(-2.19)	(-1.30)	(-2.51)	(-1.15)		(-1.28)	(0.25)	(-1.60)	(0.07)
Index Return_t		-0.362**	-0.483***	-0.218	-0.411***		-0.037	-0.023	0.027	-0.030
		(-2.35)	(-3.15)	(-1.33)	(-2.77)		(-1.55)	(-1.18)	(-0.85)	(-1.42)
MPU_t				-0.748	0.195				-0.100	0.121**
				(-1.52)	(0.37)				(-0.93)	(2.44)
GPR_t				3.221***	4.231***				0.460	0.169
				(4.10)	(4.82)				(1.34)	(0.93)
VXN				0.222**					0.066**	
				(2.39)					(2.46)	
Observations	676	676	676	676	676	676	676	676	676	676
Adjusted R-squared						0.009	0.164	0.068	-0.454	0.141
Cragg-Donald F-statistic	196.1	85.0	77.5	71.9	71.5					

Table 11: *IV Analysis of 0DTE Options Trading and Volatility[SPY]* by Brogaard, Han, and Won

Dependent variable =	First-stage estimates $0DTE\%_t$			Second-stage estimates $Ln(\sigma_t)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$0DTE\%_{t-50}$	0.570***	0.254***	0.237***			
	(6.80)	(3.47)	(3.31)			
$\widehat{0DTE\%}_t$				0.015***	0.036***	0.019***
				(4.82)	(4.67)	(2.91)
$Term\ Spread_t$		-4.220***	-3.917***		0.113***	0.043
		(-6.48)	(-6.21)		(2.68)	(1.16)
$Default\ Spread_t$		-4.944***	-5.648***		0.437***	0.293***
		(-4.75)	(-5.18)		(4.37)	(2.73)
σ_t^{forex}		13.825**	8.101*		0.049	0.030
		(2.45)	(1.66)		(0.15)	(0.12)
σ_t^{cpi}		7.926***	7.720***		-0.143	0.001
		(4.93)	(4.96)		(-1.15)	(0.01)
EPU_t		0.897*	0.155		0.075**	0.060*
		(1.72)	(0.27)		(2.09)	(1.87)
$\sigma_t^{1\,Month}$			3.114**			0.168*
			(2.49)			(1.86)
$Index\ Return_t$			-0.040			-0.093***
			(-0.29)			(-9.35)
Observations	2,970	2,924	2,924	2,924	2,924	2,924
Adjusted R-squared				0.060	0.030	0.310
Days of week FE				Yes	Yes	Yes
Cragg-Donald F-statistic	393.6	110.8	215.4			

Graph 12: Orthogonalization of 0DTE with 22-Day Rolling Volatility and VXN

```
================================================================================
SUMMARY TABLE: Comparison of All Models (Newey-West HAC Standard Errors)
================================================================================

Model 1: Orthogonalization with both 22-Rolling and VXN
Model 2: Orthogonalization with 22-Rolling
Model 3: Orthogonalization with VXN

================================================================================
Variable          Model 1              Model 2              Model 3
--------------------------------------------------------------------------------
0DTE%             -0.0004 (-0.08)       0.0032 (0.60)       -0.0008 (-0.15)
CPI               -0.0399 (-0.49)      -0.0609 (-0.74)      -0.0379 (-0.46)
Default_Spread     0.6048*** (4.72)     0.6021*** (4.72)     0.6047*** (4.72)
EPU               -0.0186 (-0.44)      -0.0158 (-0.38)      -0.0189 (-0.45)
Index_Return      -0.0463*** (-2.62)   -0.0465*** (-2.64)   -0.0462*** (-2.63)
Term_Spread        0.0117 (0.32)        0.0184 (0.51)        0.0110 (0.30)
const              1.7043*** (6.04)     1.7260*** (6.11)     1.7031*** (6.03)
--------------------------------------------------------------------------------
Observations      676                  676                  676
Adjusted R-squared 0.0733              0.0741               0.0733
Durbin-Watson     1.4177               1.4187               1.4178
```

Graph 13: Raw 0DTE Impulse Response Function (IRF)
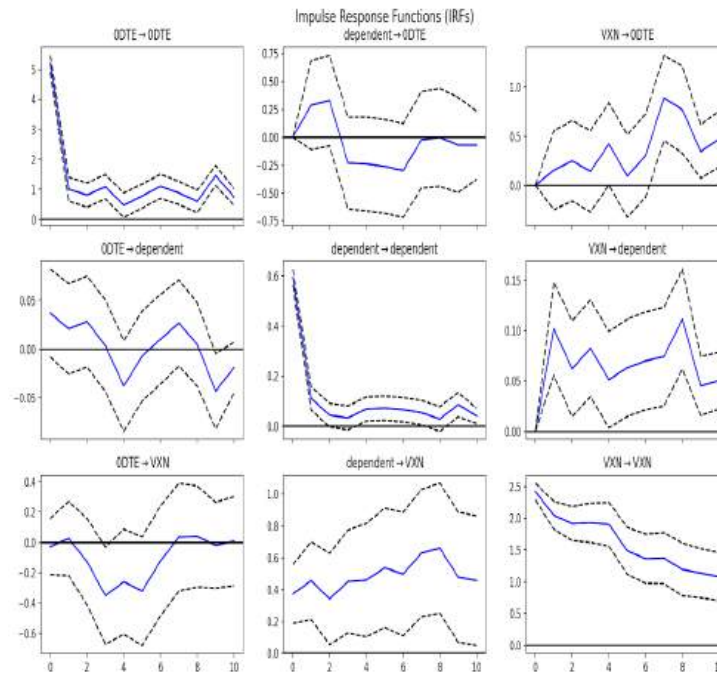
Graph 14: LAG 0DTE Impulse Response Function (IRF)
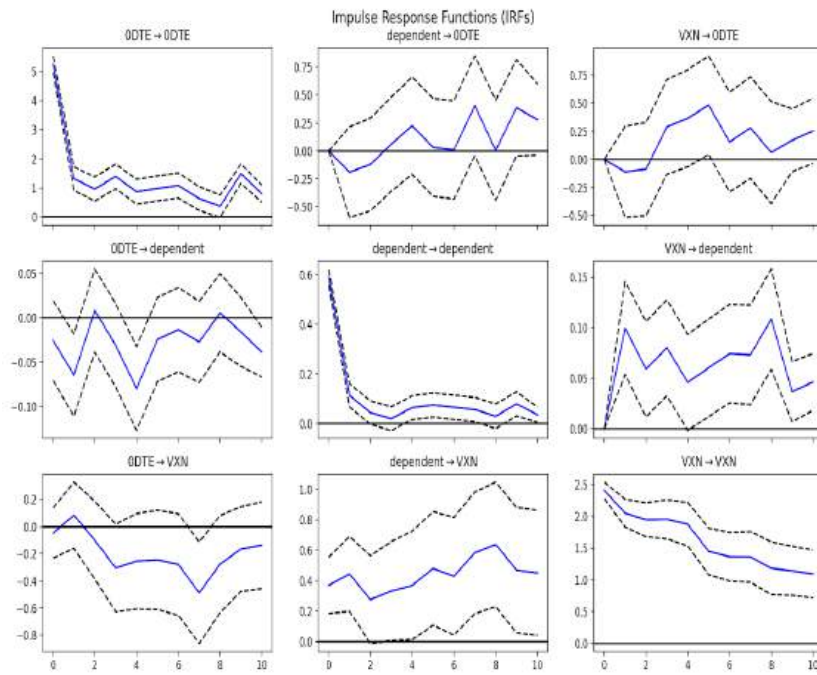
Graph 15: Orthogonalized LAG 0DTE on VXN and 22-Day Rolling Volatility Impulse
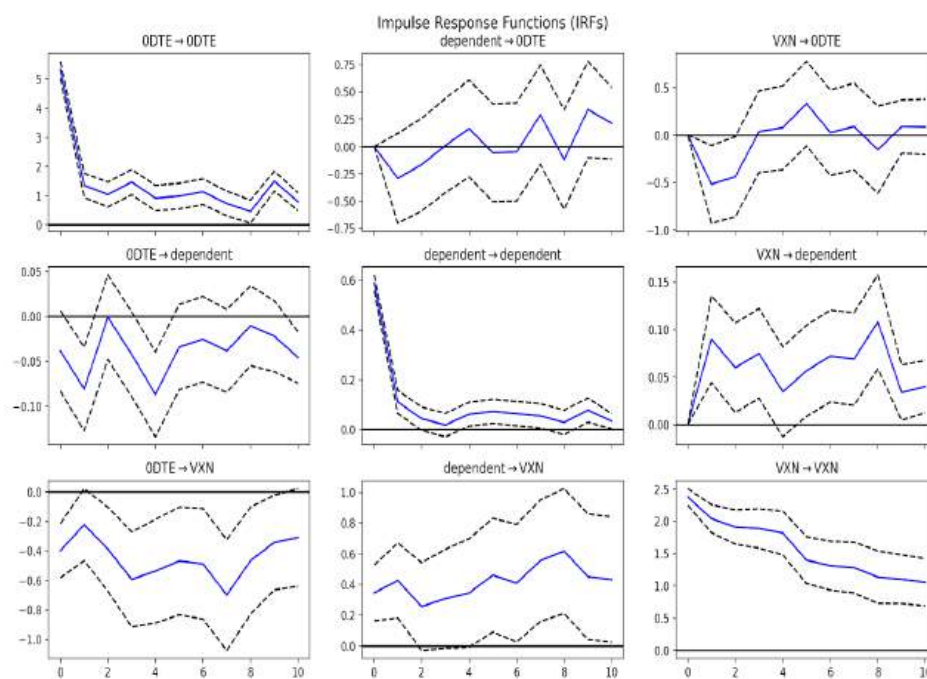Response Function (IRF)

Table 16: VAR-FEVD of Raw 0DTE, LAG 0DTE, and Orthogonalized 0DTE.

```
RAW0DTE

=== Forecast Error Variance Decomposition (FEVD) RAW 0DTE Tables ===
FEVD decomposition shape: (3, 10, 3)

--- FEVD for 0DTE ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1    100.00          0.00       0.00     100.0
1      2      0.39         99.61       0.00     100.0
2      3      0.01          2.31      97.68     100.0

--- FEVD for dependent ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1     99.62          0.29       0.08     100.0
1      2      0.48         96.78       2.74     100.0
2      3      0.01          3.33      96.65     100.0

--- FEVD for VXN ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1     99.05          0.65       0.29     100.0
1      2      0.68         95.62       3.70     100.0
2      3      0.12          3.26      96.61     100.0




=== Forecast Error Variance Decomposition (FEVD) LAG-0DTE Tables ===
FEVD decomposition shape: (3, 10, 3)

--- FEVD for 0DTE ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1    100.00          0.00       0.00     100.0
1      2      0.19         99.81       0.00     100.0
2      3      0.04          2.30      97.66     100.0

--- FEVD for dependent ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1     99.84          0.12       0.04     100.0
1      2      1.27         96.09       2.64     100.0
2      3      0.09          3.23      96.68     100.0

--- FEVD for VXN ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1     99.77          0.17       0.06     100.0
1      2      1.27         95.22       3.52     100.0
2      3      0.14          2.89      96.97     100.0




=== Forecast Error Variance Decomposition (FEVD) Orthogonalized Tables ===
FEVD decomposition shape: (3, 10, 3)

--- FEVD for 0DTE ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1    100.00          0.00       0.00     100.0
1      2      0.42         99.58       0.00     100.0
2      3      2.66          2.01      95.33     100.0

--- FEVD for dependent ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1     98.85          0.28       0.87     100.0
1      2      2.11         95.74       2.15     100.0
2      3      2.00          2.91      95.09     100.0

--- FEVD for VXN ---
    Period  0DTE (%)  dependent (%)  VXN (%)  Total (%)
0      1     98.21          0.35       1.44     100.0
1      2      2.08         94.86       3.06     100.0
2      3      2.51          2.58      94.91     100.0
```

Table 17: Robustness Appendix of VIF and VAR correlations

```
=== Combined Robustness Table (VIF + VAR Correlations) ===

          Case   Variable        VIF  Corr_0DTE  Corr_dependent   Corr_VXN
           Raw      const  33.202064        NaN             NaN        NaN
           Raw       0DTE   1.075085   1.000000        0.062132  -0.011303
           Raw  dependent   1.281616   0.062132        1.000000   0.151019
           Raw        VXN   1.366624  -0.011303        0.151019   1.000000
           Lag      const  36.795438        NaN             NaN        NaN
           Lag       0DTE   1.073701   1.000000       -0.043067  -0.019320
           Lag  dependent   1.293532  -0.043067        1.000000   0.152482
           Lag        VXN   1.374408  -0.019320        0.152482   1.000000
  Orthogonalized      const  21.699427        NaN             NaN        NaN
  Orthogonalized       0DTE   1.008381   1.000000       -0.064580  -0.163121
  Orthogonalized  dependent   1.292064  -0.064580        1.000000   0.152138
  Orthogonalized        VXN   1.283682  -0.163121        0.152138   1.000000
```

Can Your Smartwatch Peek into Your Liver? By Vansh Saxena

Abstract

Traditionally when doctors want to know how well your liver is working, they prescribe blood tests known as "Liver Function Tests," which require blood sampling and laboratory testing before they can see the results (Yetman). Now imagine walking into your primary care doctor's office for a regular check-up and being able to figure out if your liver is working well using your Apple Watch and in just few minutes. With the recent advancement in the wearable devices industry, we could now make this thought a reality.

Here is the game-changing idea: combine Photoplethysmography (PPG), an existing feature of smartwatches (Allen), with Indocyanine Green (ICG), a dye, widely used in medicine for diagnostic imaging and for checking liver health (Lu). How it can work – if a smartwatch could accurately track how quickly ICG is disappearing from the blood after an ICG injection, this would give doctors a look into liver performance, in just few minutes. This could also be an easier method as compared to traditional "Liver Function Tests," as it will not require blood sampling and waiting for laboratory testing, which in turn could take few days and laboratory resources.

How PPG and ICG can Work Together

PPG in a smartwatch works by shining near-infrared (NIR) light into the wrist and watching how blood absorbs and reflects light (Allen). Fig 1 shows the PPG concept (Timar-Fulep). As heart pumps, blood volume changes in blood vessels, which in turn increase the NIR light absorption, which creates a pulse signal (Park). Apps such as Cardiogram uses algorithms to measure time between these pulses to record heart rate (Tison). Most smartwatches use NIR light between 850 and 950 nanometers, which is a range that works well for measuring blood signals (Tech-Led). Fig 2 shows example of a smartwatch using PPG (Ernest).

ICG is a dye that has been used for many years in medical imaging and surgery (Lu). This dye absorbs NIR light very well (Lu). ICG is removed from the blood solely by the liver, usually with ICG half-life in blood of three to four minutes and ICG disappear within 15 min in healthy liver patients (Vos). For people with liver diseases such as cirrhosis, the body clears ICG slowly, thus give indication of liver dysfunction or complications after surgery (Vos).

The hypothesis presented here is that when ICG is present in blood, more NIR light will be absorbed by the blood during PPG i.e., the peak NIR light absorption signal will be higher than the regular PPG signal. If we can alter the algorithms in the smartwatch App to measure strength of the pulse signal rather than the timing between them, we can observe how long it takes for the signal from ICG to return to normal i.e., calculate how long it takes ICG to disappear from the blood. This data can help doctors to estimate how efficiently the liver is working (Vos).
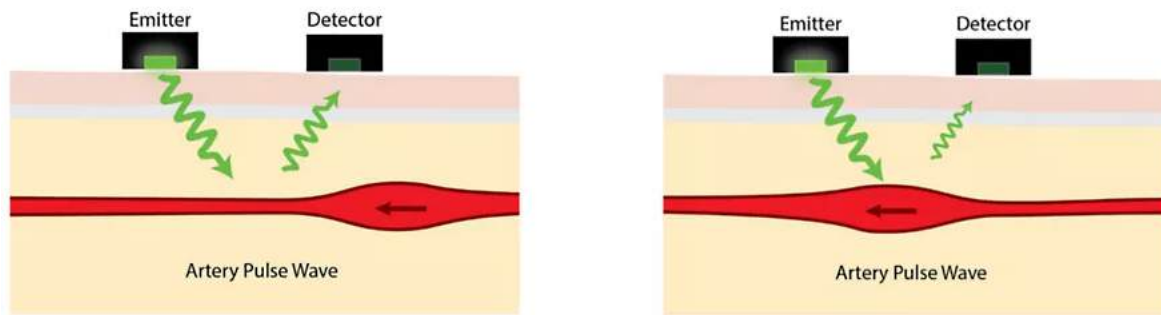
Fig. 1 Photoplethysmography Concept (Timar-Fulep).

Fig. 2 Example of a smartwatch using PPG (Ernest).

Making it Work

To make this idea work, we need to customize the existing PPG technology in wearable devices and then test this concept in a laboratory setting. The most critical step in customizing the PPG technology is to modify signal processing i.e., instead of providing timing between two pulse signals, focus on the height of the pulse signal peaks (Mejía-Mejía). Then the algorithms in Apps like Cardiogram needs to be updated as well to provide the output in form of a indicator – normal liver function or potential risk of liver disease. Once the algorithm is ready, the concept testing can be done in a laboratory. A wearable prototype can be created using affordable PPG laser and sensors e.g., PulseSensor, and commonly available dyes e.g., food dyes for testing proof of concept. A more through testing will require a prototype, which can be built using NIR laser and sensors. This prototype along with ICG can further validate the concept in laboratory. Finally, this concept needs to be tested on humans and further optimized for commercial applications.

Conclusion

Using PPG technology of smartwatches, and monitoring how quickly ICG disappear from the blood, could change how doctors check for liver health, making it possible to avoid blood tests and save time. With more research and implementation, what currently is a long procedure could soon be as simple as checking your watch.

Works Cited

Allen, John, et al. "Photoplethysmography (PPG): state-of-the-art methods and applications."
        Physiological Measurement, vol. 42, no. 10, Oct. 2021, p. 100301.
        https://doi.org/10.1088/1361-6579/ac2d82.

Ernest. "Measuring HRV: Optical (PPG) vs ECG/EKG." Medium. Last modified Oct 19, 2022.
        Accessed July 13, 2025.
        https://medium.com/terraapi/measuring-hrv-optical-ppg-vs-ecg-ekg-47194389ff6d.

Lu, Chia-Hung, and Jong-Kai Hsiao. "Indocyanine Green: An old drug with novel applications."
        Tzu Chi Medical Journal, vol. 33, no. 4, Mar. 2021, pp. 317–22.
        https://doi.org/10.4103/tcmj.tcmj_216_20.

Mejía-Mejía, Elisa, et al. "Photoplethysmography Signal Processing and Synthesis." Elsevier
        eBooks, 2021, pp. 69–146. https://doi.org/10.1016/b978-0-12-823374-0.00015-3.

Park, Junyung, et al. "Photoplethysmogram Analysis and Applications: An Integrative Review."
        Frontiers in Physiology, vol. 12, Mar. 2022, pp. 1–23.
        https://doi.org/10.3389/fphys.2021.808451.

Tech-Led. "How Near Infrared LEDs Are Revolutionizing Medical Devices." Marubeni.
        Accessed July 13, 2025.
        https://tech-led.com/how-near-infrared-leds-are-revolutionizing-medical-devices.

Timar-Fulep, Csilla. "Using Photoplethysmography (PPG) for Designing Optical Heart Rate
        Sensors." Ansys. Last modified June 27, 2023. Accessed July 13, 2025.
        https://www.ansys.com/blog/modeling-human-skin-and-optical-heart-rate-sensors.

Tison, Geoffrey H., et al. "Passive Detection of Atrial Fibrillation Using a Commercially
        Available Smartwatch." JAMA Cardiology, vol. 3, no. 5, Mar. 2018, pp. 409–416.
        https://doi.org/10.1001/jamacardio.2018.0136.

Vos, J. J., et al. "Green Light for Liver Function Monitoring Using Indocyanine Green? An
        Overview of Current Clinical Applications." Anaesthesia, vol. 69, no. 12, June 2014, pp.
        1364–76. https://doi.org/10.1111/anae.12755.

Yetman, Daniel. "Liver Function Tests: Purpose and Procedure." Healthline. Last modified
        March 21, 2023. Accessed July 13, 2025.
        https://www.healthline.com/health/liver-function-tests.

Serial Position Effect: Among High School Students of India By Akshara Kejriwal

Abstract

The serial position effect is a significant effect that deals with abilities of memory recall depending on the position of the information. The main components are primacy, where individuals tend to remember the first few items, and recency, where individuals tend to remember the last few items of the sequence or list. The objective of this study is to explore the serial position effect's validity when the type of words differs from one another. Ten participants were examined to conduct this study. All of them were high school kids who belonged to India. All 10 participants were given 3 different lists with different word types —randomly chosen, rhyming, and connected words. Each word was read aloud by the researcher within 1.2 seconds, and after the word list was read out, the participants were instructed to write their answers on the provided answer sheet to record their responses. The procedure continued for the remaining lists of words as well. The findings revealed that the effect was partially true for the three lists, and the results drastically differed from one another, which led to an insightful discussion. In conclusion, the study highlights primacy and recency in memory recall while acknowledging contextual limitations and the need for further research.

Keywords: Serial Position Effect, Memory, Primacy, Recency, high school

Introduction

Memory is the faculty of the mind that helps decode, store, and retrieve information for our future actions. The human mind is complex to understand, which makes memory even harder to understand (Robertson, 2002). Studies related to memory and research focus on the ability to retrieve stored information by means of free recall. Memory is not a single process but a combination of varied systems—such as sensory memory, short-term memory, and long-term memory—each having its own functions and limitations **(Brady et al., 2008**). One way in which the connection can be witnessed is through a universal effect, which is the serial position effect. It is a psychological phenomenon where people tend to remember the first and last few words in a list or sequence instead of the middle words. This effect was recognised by Hermann Ebbinghaus. He discovered the Serial Position Effect (SPE) in the late 19th century, which enables us to understand how information in a sequence affects memory patterns (Jensen, 1962). The serial position effect comprises of two main components, which are primacy and recency, which help individuals recognise this effect better. Primacy is a cognitive bias where people tend to remember the first few words of the list or the sequence (Li, 2010). While recency can be defined as the opposite of primacy, where people tend to remember the last few words of the list or the sequence. Subsequent studies and in-depth research indicate that the SPEs are well known for examining memory in tasks such as free recall, serial recall, recognition memory, and implicit memory (Cortis Mack et al., 2017). The Atkinson-Shiffrin multi-store model recognises the connection of short-term and

long-term memory with primacy and recency. It states that the individuals who tend to remember the first few words have a long-term memory, while those who only remember the last few words have a short-term memory.

Understanding the SPE has important practical implications in various aspects of our everyday life, from learning and education to marketing and advertising (Dunlosky et al., 2013). Though it is termed as a universal effect, **states** that many confounding variables can hinder this effect. The consideration of these variables needs to be considered for the credibility of this effect. This study focuses on one such variable in which the type of words presented in the list or sequence influences the participant's recall abilities. The nature of words presented in the list or sequence given can have an influence on the shape of the curve; different categories of words exert distinct results on memory performance, indirectly resulting in the varying degree of recall. There have been quite a few studies done on this topic, but it is unexplored, especially in the cultural context. India serves as an appropriate context for exploration due to its immense diversity explored in individuals, which is needed for the ecological validity. The need for this study is significant in proving the generalisability of this effect, highlighting the universality of it within diverse individuals.

Literature Review

The serial position effect describes the tendency to remember the first and last items in a series more accurately than those in the classic U-shaped curve. This discovery, initiated by Ebbinghaus (1885), first noted this pattern, and later subsequent research linked this effect with the multi-store model of memory (Atkinson & Shiffrin, 1968). This link between the effect and the multi-store memory stated that primary words were associated with long-term memory (Rundus, 1971; Atkinson & Shiffrin, 1968). While the later-mentioned words do not require much recall, which is probable to be encoded in the short-term memory, the recency effect (Glanzer & Cunitz, 1966; Baddeley & Hitch). Empirical support comes from studies of healthy participants and brain-damaged patients, both of which show primacy and recency advantages (Glanzer & Cunitz, 1966; Baddeley & Hitch, 1974). Exploring this effect further, few researchers discovered the influence of the word type given, which directly affects this effect (Berger,2016). There have been few studies on how the word type can weaken or strengthen this pattern and contribute to the U-shaped curve (Snell, and Grainger,2019). Studies dealing with abstract and concrete word lists show that concrete items typically benefit from richer recall abilities, enhancing primacy. While the abstract words, lacking vivid imagery, produce flatter recall curves, suggesting that reduced semantic elaboration diminishes the primacy advantage (Paivio, 1971). When the list consists of words that are randomly selected or sound familiar, then the serial position effect tends to remain universally valid. Emotionally salient words (especially negative) are recalled more often at both early and late list positions, amplifying both primacy and recency. Another well-studied phenomenon, termed temporal clustering, refers to participants' tendencies to successively recall items that occupied neighbouring positions in the studied lists (Kahana, 1996). Murdock (1962)

replicated the SPE with randomly ordered words and showed that without meaningful associations, participants rely primarily on short-term memory processes, producing a clear U-shaped recall curve. In contrast, the words that are phonologically similar tend to blur together, which makes it hard to distinguish for the recency effect, which is linked to the short-term memory. Subsequent studies indicate that phonological similarity can negatively influence the short-term memory. Baddeley (1966) discovered that lists of acoustically similar words were recalled less accurately than acoustically distinct words, suggesting interference at the encoding stage. However, rhyme can also serve as a retrieval aid, as demonstrated in studies of rhyme-based mnemonics (Nelson & McEvoy, 1979). Lastly, the words that are connected to each other and share a link between each other show a negative response to the serial position effect (**Li, and Epley, 2**009). When items share relationships or form coherent units, memory performance improves, particularly for middle-list items that usually suffer from poor recall (Bower & Clark, 1969). In view of this, more empirical studies are needed to ascertain the consistency of this SPE theory across different contexts, participants, and variables.

Research Methodology

Research Design
　　　This study employs a repeated measure experimental design which increases the statistical power by controlling individual differences (Perugini, Gallucci, and Costantini 2018). The participants are exposed to three types of word lists, which are random, similar, and rhyming, containing words. The number of subjects is 10 high school students. This study gives us direct information on the influence of the type of word in a list or sequence on the memory abilities of the participant. The independent variable (IV) is the word type, and the dependent variable (DV) is the number of words correctly recalled from each list.

Participants
　　　The sample consists of 10 high school students (aged 15-18 years) who were recruited through purposive sampling which provides in-depth understanding, cost-effectiveness, and efficiency for small-scale studies (Campbell et al.,2020). All participants were proficient in English and had no reported learning or memory disabilities. The small sample size is chosen for feasibility. The study was conducted in a classroom setting.

Materials
　　　Three sets of 14 words were given, where each set had a certain type of words.

1. Random words—these words are not linked to each other.
2. Rhyming words—these words have rhyming words in a certain order.
3. Connected words—these words are linked to each other, which conveys a story through words.

The instructions, along with the words, were read aloud to the participants. They were provided with a pen and paper to note down the words. A timer was also used appropriately to maintain time duration and ensure equal representation.

Procedure

Participants were seated in a quiet classroom to prevent external variables that might affect the results. The researcher read the first 14 words aloud at a rate of one word within two seconds. After the completion of the first set, the participants were given one minute to recall and write as many words as they could remember in any order. The above procedure for each list was carried out two more times and was counterbalanced to prevent order effects. The type and number of these words were noted from each set, which was recorded as the dependent measure.

Ethical Considerations

The informed consent was taken by participants prior to the procedure. Participants were briefed on the instructions and the purpose of the study but were not told about the hypothesis. There was assurance of information being confidential throughout the study. Participation was voluntary, and the participants had the right to withdraw from the experiment anytime during the experiment. Lastly, a session was conducted to explain to the participants the serial position effect and the purpose of the research.

Data Collection Procedure

All 10 participants participated in this experiment of the serial position effect that tested their recall abilities. There were three lists given, each list containing 14 words of its own type – randomly chosen, rhyming and connected words. The participants were instructed before the experiment was conducted by the researcher. The researcher read each word out till the fourteenth word and commanded, 'Go!' The participants were instructed to write the words on the answer sheet provided in any order they could recall. The same procedure was used for the remaining lists. There were 6 steps in this procedure, which are as follows:

Step 1 – Request the participant to sit in a seat which they felt comfortable in, which took place in a classroom setting.
Step 2 – Respond to any questions the participant might have before the experiment.
Step 3- Given the instructions required before the experiment to conduct the experiment smoothly.
Step 4 – When the participant is comfortable and prepared, start by reading the first list and say Go after the list is read aloud and tell them to write as many words as they can recall.
Step 5– Step 4 will be repeated for lists 2 and 3.
Step 6- The above steps will be repeated for all the remaining 9 participants.
The answer sheet—

S.No.	List 1	S.No.	List 2	S.No.	List 3
1	River	1	At	1	Sun
2	Bat	2	Bat	2	Brush
3	Paint	3	Cat	3	Shower
4	Mug	4	Chat	4	Breakfast
5	Door	5	Fat	5	School
6	Pen	6	Gat	6	Lunch
7	Key	7	Flat	7	Disperse
8	Ears	8	Brat	8	Home
9	Hair	9	Mat	9	Homework
10	Glasses	10	Rat	10	Play
11	Clock	11	Sat	11	Dinner
12	Grass	12	Hat	12	Nightsuit
13	Pillow	13	That	13	Lights
14	Flag	14	Spat	14	Sleep

The scoresheet 1 (Table 1) (randomly picked words)

S.no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total No.
A															
B															
C															

D															
E															
F															
G															
H															
I															
J															
Total															

Table 2 (Rhyming words)

S.no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total No.
A															
B															
C															
D															
E															
F															
G															
H															
I															
J															
Total															

Table 3 (Connect words)

S.no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total No.
A															
B															
C															
D															
E															
F															
G															
H															
I															
J															
Total															

<u>Limitations</u>

The first limitation of the study is the size of the sample, where only 10 students are recruited for the research. With such a limited group, the generalisability of the study reduces, and the lack of space for individual differences is disproportionately distributed. Another limitation is the purposive sampling, where the random participation in a larger population is harder to obtain, and it reduces the equal chance of participation from the target population. This restricts ecological validity because the confounding variables were not controlled, and the results may not accurately represent memory abilities considering the diversity in different contexts.

Data Analysis

The data analysis was conducted immediately after the experiment. The steps are as follows:

Step 1 – The response. Three similar scoresheets were designed to record the specific number of words recalled from each list. The response of the subject will be determined by this scoresheet. If the subject recalls the first word correctly from the first list, then the first column, which is assigned for the first word, will be marked with a "Yes". If the participant misspells a word or leaves a blank, then that designated place will be marked as a "No".

Step 2 – Recording the Correct Response. Since the word lists vary from one another, the participant's total no. of words correctly recalled will be recorded in the last row of each column as provided in Table 2. If the participant correctly answers 10 words, then it will be noted in the last row of the table. This procedure will repeat for the second and third lists as well.

Step 3 – Inserting the data. The scoresheet is provided to record the response of the participants. The purpose of the scoresheet is to keep data organised, obtaining the precise data of each participant systematically. The columns are from 1 to 14, and the rows are arranged for each subject. Insert the number of correct responses by the participant into each of the columns, respectively.

Step 4 – Adding up the Correct Response. In the same scoresheet, after adding all the values, add all the correct answers for each of the serial positions. The maximum number of correct responses which can be obtained is 30 (10 subjects and 3 lists).

Step 5 – Analysis of the Serial Position Effect. The serial position curve from the study was compared with the idealised curve to determine whether they exhibit a similar pattern. Analyse whether they are similar or differ from each other
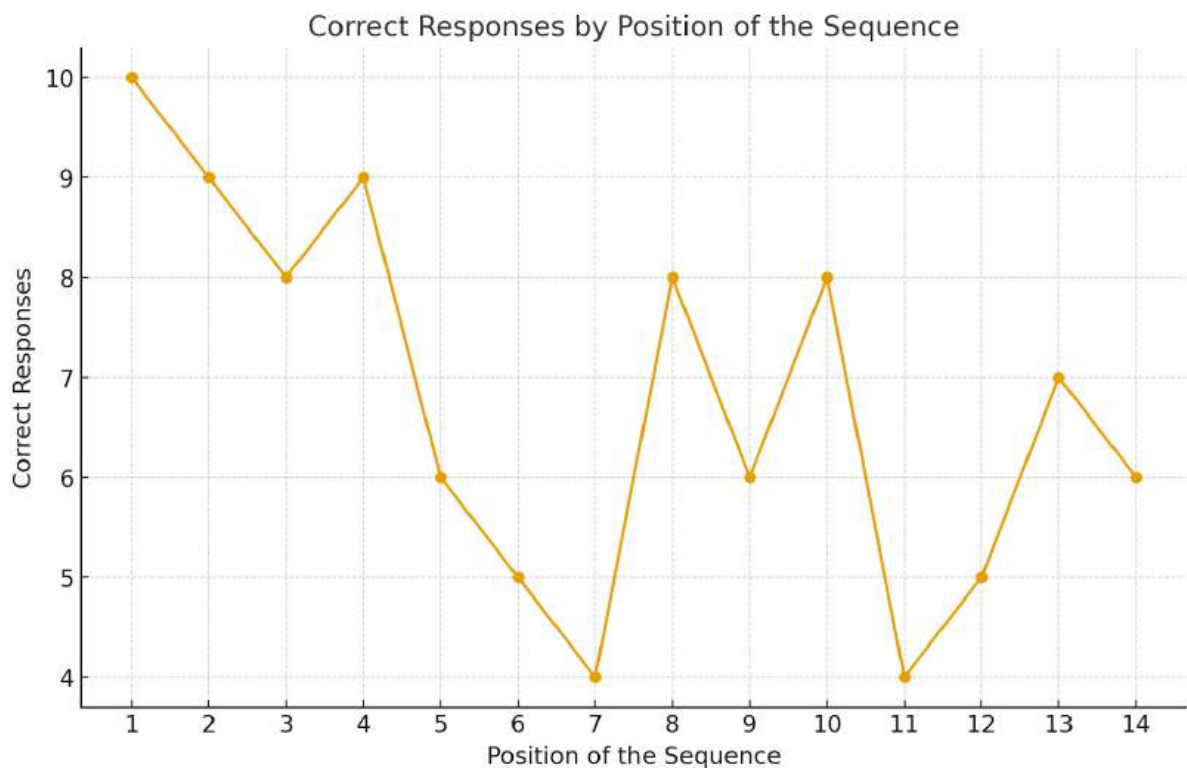
Figure 1: The correct responses of randomly picked words

In Figure 1, it is visible that the highest recalled words are from positions 1 to 4, which belong to the primary words of the sequence. There is a significant decline in the curve, which reaches its lowest at the 7th position. The rise is observed after the 7th position, which remains constant for the 8th and 9th positions and reaches a lower position for the last position. This pattern suggests that performance is initially strong but severely impaired toward the middle of the sequence.
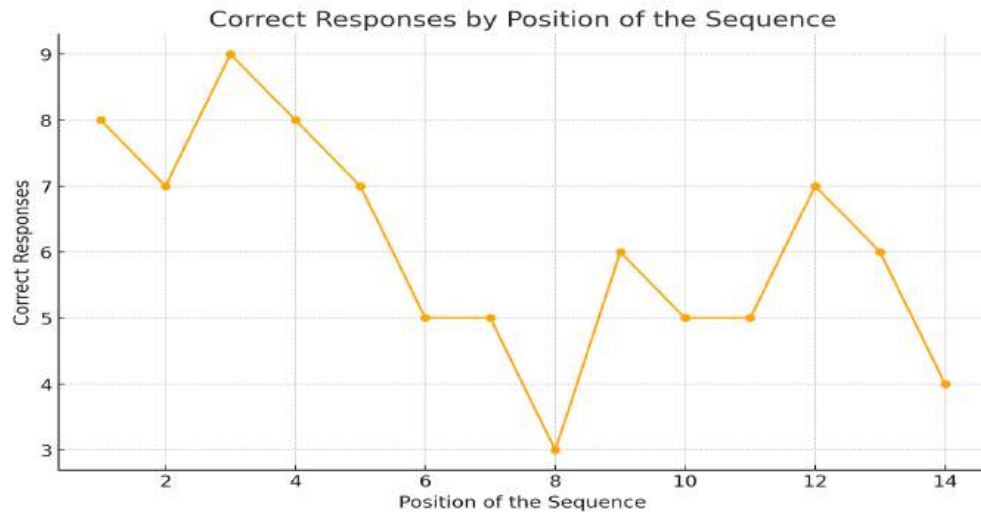

Figure 2: The correct responses of rhyming words

Figure 2, which represents the curve for rhyming words, is contrasting as compared to Figure 1. The curve remains highest in the first-positioned words belonging to the primary words in the sequence. This curve resembles the ideal serial position curve, where the highest point of the curve belongs to the primary and recency words (Deese,1957). The lowest decline of the curve is observed in the 8th position, which was the least recalled. This overall pattern highlights early retention strength, significant loss in the mid-sequence, and unstable, fluctuating performance thereafter.
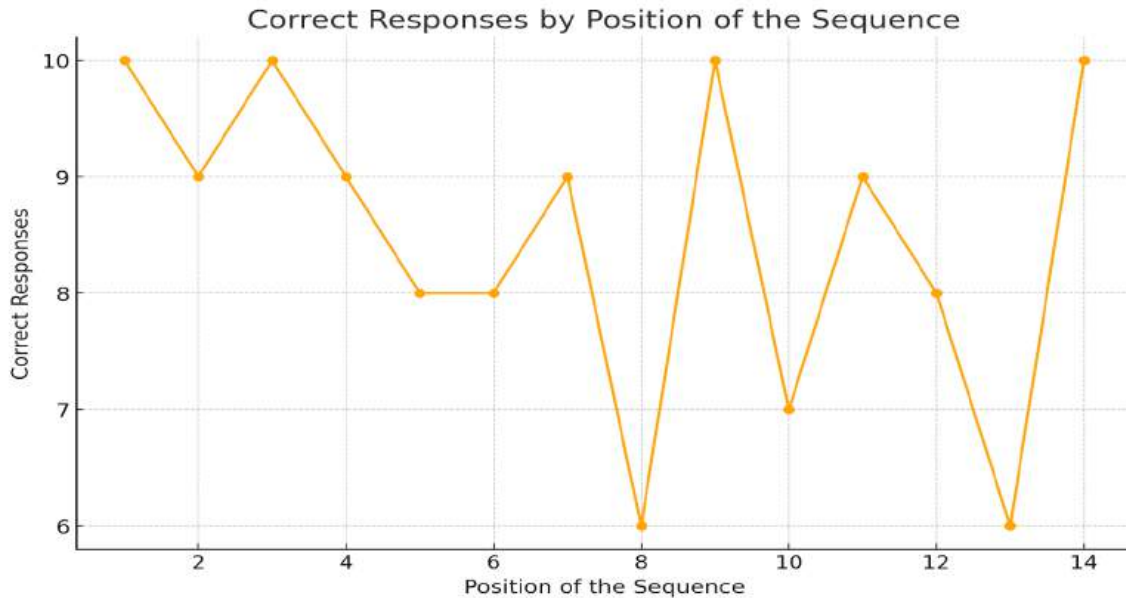
Figure 3: The Correct Responses of Connected Words

In in Figure 3, which represents the curve for connected words, the curve observed is extremely volatile with an erratic pattern. The performance starts at a peak of 10 (Position 1), drops slightly, and then reaches another peak of 10 at Position 3, demonstrating strong primacy effects. Following the curve, there is a fluctuating slope from a decline to an increase. The curve reaches its lowest point at the 8th position, like Figure 2. The curve continues to be fluctuating, reaching the highest point in the last positioned word, which suggests a very strong, late-occurring recency effect.
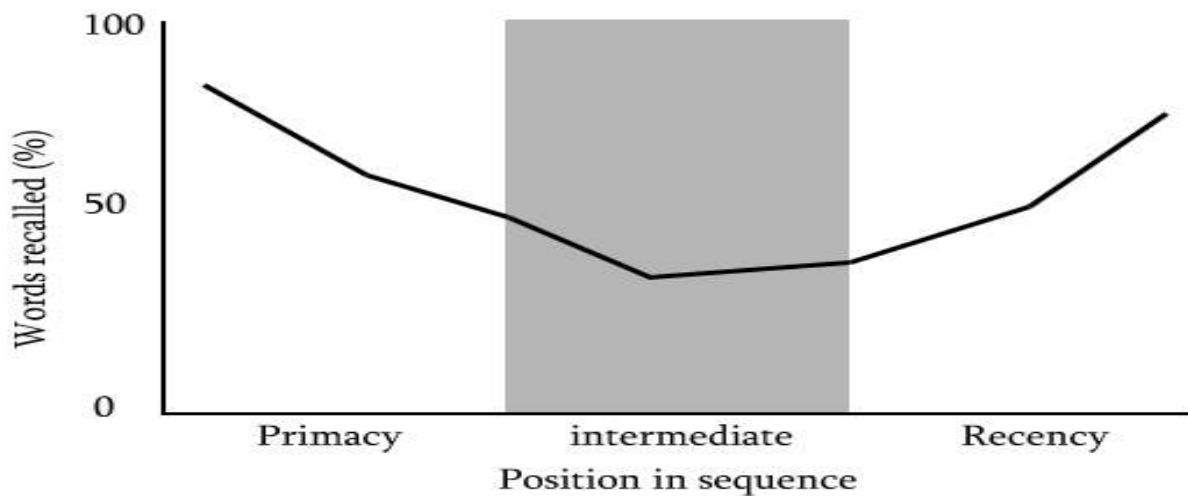
Figure 4: Serial Position Effect (Glanzer and Cunitz, 1966)

Figures 1, 2, and 3 are graphs showing the serial position effect curve of each of the word type sequences provided to the participants. The curve represents the change in the number of correct responses as the position in the sequence increases. It explores the factors of primacy and recency, which are contributing to the slope of the curve. Figure 4 is the appropriate representation of the serial position effect, where the highest recalled words belong to the primacy and recency categories.

Discussion

The study examined the effect of the serial position effect on high school students when provided with different types of words in a sequence. The results of this research provide clear evidence that the serial position effect is influenced by the type of word list presented—specifically, whether the words rhyme, have meaningful connections, or are randomly selected have an influence on this effect.

 Words that sound more like each other than lists of dissimilar words are a cornerstone of the literature on short-term memory (STM) that has shaped theories of memory for over half a century. The acid bath hypothesis proposes that the retention interval (the time between learning and recall) is affected by the number of similar items stored in memory, which is associated with the rhyming words (Posner, M. I., & Konick, A. F., 1966). Conrad and Hull (1964) first described it by comparing the recall of acoustically similar and dissimilar letters; it was quickly determined that the effect of similarity was to disturb the order of recall in the serial recall task, rather than the likelihood of retrieving the items at some point during recall (c). Words when randomly picked are assumed to produce an ideal curve of serial position effect, with primary and recency words having higher recall when compared to the intermediate words. The result of the experiment by Ebbinghaus (1885) is not completely valid for these high school students in the present study, as the words of primary and intermediate words emerge to be highly recalled when compared with recency words, which proves the experiment partially true. This finding also partially contrasts with many past findings (e.g., Murdock, 1962; Bireta et al., 2018; Lee et al., 2019; Koo et al., 2021) that support the SPE theory. Connected words enhance the serial position effect, with highly connected words (high-degree centrality) being recalled better across all positions, and this benefit is diminished when high-degree words are interspersed with low-degree words in a list. Related items produced a stronger primacy advantage because semantic similarity boosted encoding; the temporal order of presentation was the main driver of the effect (Was, Woltz, and Hirsch, 2018). The results of the present study support previous research on connected words, where the graph shows the disturbance in the serial position effect where all three types of positions are seen to be recalled. Words when picked randomly is expected to have a higher possibility in validating the Serial Position effect but the results of the study prove it otherwise, with higher recall of primary and intermediate words rather than recency words.

Conclusion

The present study proves that serial position is not completely true when provided with various types of words in ten high school students from India. Nevertheless, this study has contributed to finding a universal state of the serial position effect with a different context and consideration of a confounding variable, increasing a deeper understanding about the effect. The study did enhance the understanding of a human's recall abilities with a cognitive perspective with the help of the phenomenon of serial position effect. The major roles of primacy and recency were clearly observed in this study. This gave an insight into the position of the data having a major impact on an individual's recall abilities and retrieval of information. The link of the primacy effect to long-term storage via rehearsal and the recency effect to short-term working memory offers a foundational tool for researchers and clinicians to study, diagnose, and understand memory function and impairment. The present study has its own limitations and cannot be generalised because of the number of participants from different contexts. This study can be conducted by other researchers in different contexts with considerations of external variables to determine the credibility of the effect on individuals.

Works Cited

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval?. *Memory & cognition, 21*(2), 146-155.

Baddeley, A. D., & Hitch, G. J. (2022). Recency reexamined. In *Attention and performance VI* (pp. 647-667). Routledge.

Berger, J. (2016). *Invisible influence: The hidden forces that shape behavior*. Simon and Schuster.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior, 8*(3), 323-343.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325-14329.

Craik, F. I., & Watkins, M. J. (2016). Short-Term Memory[1]. In *Memory, Attention, and Aging: Selected Works of Fergus IM Craik* (p. 36).

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology, 54*(3), 180.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58.

Jensen, A. R. (1962). Spelling errors and the serial-position effect. *Journal of Educational Psychology, 53*(3), 105.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & cognition, 24*(1), 103-109.

Li, C. (2010). Primacy effect or recency effect? A long‑term memory test of Super Bowl commercials. *Journal of Consumer Behaviour: An International Research Review, 9*(1), 32-44.

Li, Y. E., & Epley, N. (2009). When the best appears to be saved for last: Serial position effects on choice. *Journal of Behavioral Decision Making, 22*(4), 378-389.

Mack, C. C., Cinel, C., Davies, N., Harding, M., & Ward, G. (2017). Serial position, output order, and list length effects for words presented on smartphones over very long intervals. *Journal of Memory and Language, 97,* 61-80.

Nelson, D. L., & McEvoy, C. L. (1979). Effects of retention interval and modality on sensory and semantic trace information. *Memory & cognition, 7*(4), 257-262.

Paivio, A., & Csapo, K. (1971). Short-term sequential memory for pictures and words. *Psychonomic Science, 24*(2), 50-51.

Perugini, M., Gallucci, M., & Costantini, G. (2018). A practical primer to power analysis for simple experimental designs. *Revue Internationale de Psychologie Sociale, 31*(1), 1-23

Posner, M. I., & Konick, A. F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology, 72*(2), 221.

Robertson, J. (2002). Diamond-like amorphous carbon. *Materials Science and Engineering: R. Reports, 37*(4-6), 129-281.

Snell, J., & Grainger, J. (2019). Word position coding in reading is noisy. *Psychonomic Bulletin & Review, 26*(2), 609-615.

Was, C., Woltz, D., & Hirsch, D. (2019). Memory processes underlying long-term semantic priming. *Memory & cognition, 47*(2), 313-325.

Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *The American Journal of Psychology, 78*(4), 567-574.

The impact of digital learning on adolescent social development during the Covid-19 pandemic By Yuvraaj Sheth

Abstract:

The Covid-19 pandemic spread rapidly in May 2020 affecting 1.6 billion children. Schools were forced to lock down which resulted in the rise of digital learning. As this was the first instance in which all students had to rapidly adapt to digital learning, it is unclear how digital learning has impacted social development. This review investigates the question, "How has digital learning during the Covid-19 pandemic affected the development of social skills among adolescents?" This paper provides insights on the effects of the Covid-19 pandemic and digital learning on the social development and mental health of the students. Research studies across the globe support the idea that adolescents experienced high levels of depression and anxiety throughout the pandemic. However, limitations, due to the narrow time frame of the study, restricted data collection. A social learning theory presents the idea that people learn social skills by observing one another. The deprivation of direct observation and social cues in online learning led to poor social skills of adolescents. These poor social skills may have led to mental disorders such as depression and anxiety. With the negative effects of the Covid-19 pandemic established, some strategies to mitigate future scenarios of digital learning would be to incorporate social-emotional learning into teaching.

Introduction:

The Coronavirus Disease 2019 (Covid-19) pandemic had a profound impact on over 1.6 billion children worldwide (New global tracker to measure pandemic's impact on education worldwide, 2021). As physical learning was shut down due to the spread of the virus, digital learning became the primary alternative. Now that this learning style has been incorporated into our society, one must understand how it has affected the adolescents of today. To understand how digital learning during the Covid-19 pandemic affected the development of social skills among adolescents, we must first examine what caused digital learning itself. The Covid-19 pandemic spread rapidly in May 2020. This led to the closure of schools due to concerns about contact in such a populated environment. School closures introduced the issue about how and where the students would continue learning. As a result, digital learning became the primary solution to this dilemma (Depression (major depressive disorder), 2022; Social Skills, 2022). While digital learning gave a necessary solution to the education continuity, its long-term effects on adolescent social development remain understudied due to the pandemic's recency. This review investigates the question, "How has digital learning during the Covid-19 pandemic affected the development of social skills among adolescents?" This paper will shed light on the effects of the Covid-19 pandemic and digital learning on the social development and mental health of the students of today. It will also provide insights for how these obstacles can be overcome in the future.

The Covid-19 Pandemic

According to the Center for Disease Control and Prevention (CDC), on January 10th, 2020, a group of patients in China's Hubei Province started experiencing symptoms of a pneumonia-like illness that was unaffected by standard treatments. This virus was known as Covid-19 (CDC Museum COVID-19 Timeline, 2023). According to the National Cancer Institute, Covid-19 is a virus that causes a respiratory disease which can spread from person to person through droplets released when an infected person coughs, sneezes, or talks (Definition of COVID-19, 2020). It also may spread through touching a surface with the virus on it and then touching one's mouth, nose or eyes. This contributes to how easy it was for the disease to spread from person to person. In just a few weeks, Wuhan, China–a city of 11 million people–was placed under a lockdown due to the virus (CDC Museum COVID-19 Timeline, 2023). Just a day later, the CDC confirmed a case of a travel-related infection of the Covid-19 in Washington State. This was the instigator that set off the explosion that led to the rapid spread of the pandemic, which ultimately affected billions. Just 3 months later on April 10th, with 500,000 cases and over 18,000 dead, the United States became the country with the most reported Covid-19 cases and deaths. Other countries were experiencing similar problems with the pandemic. For example, in the UK, 1 in 160 individuals had the Covid-19 virus (Coronavirus (COVID-19) roundup, 2025). Infection rates were the highest in ages 16-24, half of which were adolescents.

People in these countries that had disparities in income were forced to stay inside without any access to technology. This is known as the Digital Divide and due to this issue, many students were unable to continue their learning. This problem was apparent in India with rural areas. For example, Bihar had some of the lowest numbers of internet users with only 32.45 per 100 people having access to the internet (Krishnan, 2020). In contrast, areas such as Delhi, Himachal Pradesh, Kerala, and Maharashtra were among the top five areas in terms of internet users relative to the rest of the country. Other areas of the globe were similarly impacted. Furthermore, according to Unicef, 247 million students enrolled in elementary and high schools had to stay home due to concerns about the virus spreading in India (UNICEF. COVID-19: Schools for more than 168 million children globally have been completely closed for almost a full year, says UNICEF, 2021). Numerous countries were forced into lockdowns, preventing students from continuing their in-person education. This was where the primary solution was introduced, digital learning.

Digital Learning

Digital learning is any type of learning that is driven primarily through technology. This includes platforms such as online classes, virtual chat boxes, and educational apps to help enrich learning for youth. Specifically, online classes were conducted over a variety of platforms including Microsoft teams, Google classroom, and most prominently, Zoom. Virtual platforms were used as a way to provide the essential resources that students would regularly receive in person. Blackboard, Moodle, and primarily Canvas were used as a

learning management system for the instructor in a college based setting (Choosing an LMS: Canvas vs Moodle vs Blackboard vs LMS365, 2024). Without these platforms, there would be no reliable source for submitting assignments and an online grading system. Finally, educational apps were used as supplementary aid for a variety of subjects, such as Khan Academy for math, Quizlet for studying, and Duolingo for language learning. Together, all of these online options replaced physical learning. These adaptations were a necessary and rapid response to the pandemic in order to ensure the continuity of education for adolescents. With investments from 2019 prior to Covid-19 reaching $18.66 billion in digital learning technology, there was already a strong argument for why digital learning should replace in-person school (What is Digital Learning? 10 Benefits of Digital Learning, 2024). Even though some schools still used analog tools before Covid-19 such as notebooks, whiteboards, and paper textbooks, newer universities and schools had tried experimenting with newer technology, including iPads or Chromebooks for online learning websites and apps. Although the benefits of digital learning in a pandemic period were vast, the cons of digital learning impacting student social skills is prevalent.

Development of Social Skills

Social skills are the abilities we use to interact with one another effectively. These skills can be both non-verbal and verbal including speaking communication, body language, facial expressions, and gestures. Social skills are developed through complex synaptic processes at the neurological and psychological level (Soto-Icaza et. al., 2015). Neurotransmitters such as oxytocin and serotonin play a crucial role in mood regulation and social bonding. Oxytocin is prevalent in processes that involve social bonding and trust, while serotonin has a prominent impact on mood and social behavior.

Traditional behavioral theories, such as one proposed by Ivan Pavlov's Classical Conditioning, are learning processes in which an automatic, conditioned response is paired with specific stimuli (Rehman et. al., 2024). A theory by Albert Bandura called the Social Learning Theory presents the idea that people learn social skills by observing one another (How Social Learning Theory Works, 2024). This theory bridges the wide gap between traditional behavioral theories which support direct reinforcement, and cognitive theories, which focus more on internal processes and stimuli. In the context of social learning, supporting direct reinforcement means that when someone learns a behavior by seeing others, this behavior can be reinforced through reward and/or punishment. For example, a child could observe another child

share their things and get praised which would be positive reinforcement for the child that was observing. On the other hand, a child who observes another child who gets scolded for whining will be an example of negative reinforcement. These are very primitive types of direct reinforcement, but they still apply to a larger scale of social skills. In Albert Bandura's book, Social Learning Theory, he states that most human behavior is learned observationally primarily through modeling, which is observing others to develop an idea of how new

behaviors are performed, and this information is used as a guide for action (How Social Learning Theory Works, 2024).

Adolescence is one of the most sensitive and critical moments of social development as it is characterized by hormonal and biological changes that occur through puberty (Orben et. al., 2020). When a child enters adolescence they are exposed to new environments such as middle school and high school. The formation of complex peer relationships also grows during this time (Lam et. al., 2014). Simultaneously, the development of cognitive abilities such as critical thinking, executive function control, and self-referential processing emerge as puberty progresses. This development helps adolescents reflect upon themselves, thus leading to better navigation of social networks (Orben et. al., 2020). The brain itself is easy to change and manipulate during these crucial moments of adolescence. This plasticity is mainly driven by the processes of neurogenesis and synaptogenesis. Neurogenesis is the formation of new neurons in the brain. While this is most active during infancy and early child development, it continues into adolescents and more scarcely into adulthood. In adolescence, neurogenesis plays a role in a variety of factors such as memory, emotional regulation, and learning. These three roles are especially crucial in adolescent development. Synaptogenesis is the creation of synaptic connections between neurons. During adolescence, the brain undergoes immense amounts of synaptogenesis, especially in areas in the prefrontal cortex, which is responsible for decision-making, impulse control, and planning. Again, this is especially important for adolescents because the development of these brain areas will determine their social, educational, and personal capabilities. The brain refines its connections through a process known as synaptic pruning. In this process, the brain disposes of underused synaptic connections to strengthen existing ones which results in a refined and optimized usage of synapses (Sakai, 2020). Although this process primarily occurs in early childhood, it continues well into adolescence.

Social Cues

The definition of social cues are the verbal or non-verbal signals expressed through the body, voice, face, and motion which impact conversations by influencing the tone and impressions of how we convey a message. When we see a face, we also sense the emotion behind it, which includes what someone is trying to convey. In such a small space, a vast amount of stimulus is presented. With all of these tiny details, we piece together a representation of what the person is trying to communicate (Adams et. al., 2017). Without these details, a large gap is formed in our brain, which is left for interpretation. The context clues are missing, resulting in a lack of proper communication. Thus, this "social vision" implies the fact that our brains are tuned into the social cues which are provided from in-person communication. Our brain's neocortex, the most advanced part of our brain responsible for consciousness and judgment, is also responsible for interpreting socio-visual cues presented by facial expressions.

Immediate Impacts of Digital Learning

The most immediate social impact of digital learning is that it can create isolation. In an article by Stevenson University, they explain how one of the biggest factors in communication is using nonverbal communication such as body language (The Importance of Effective Communication, 2025). Without body language, it is difficult to interpret the overall tone of what one is saying without the presence of any sort of visual cues. Since it is difficult for a student to effectively communicate with their peers during digital learning, it will ultimately limit social interaction. This type of communication is not effective in fostering social connections. A research study conducted by Cingel et al., 2022 collected data from 1,256 United States adolescents with ages ranging from 14 to 16 and found that teens who were able to attend school in-person felt more inclusion in their social groups as opposed to those in online learning. They also reported that social media was not a sufficient aid in supplementing these social interactions (Cingel et. al., 2022) . In this same research article, they used a diathesis-stress model to predict that only certain individuals would have been more likely to experience the mental health disturbances based on individual factors. They theorized that the causation by the pandemic was prevalent the most among adolescents who were in online learning. The group conducted an online survey of people being the ages of 14-16 from May 20, 2021 through June 23, 2021. In the survey, they asked participants how they participated in the school during the month. The responses included "in-person in the school building", "hybrid model where some school is in-person and some is virtual/online", "completely virtual/online", "did not attend school at all", or "a different variation of schooling." These options helped determine the differences between each category. If there was not any separation between categories then it would've led to a lackluster finding. They used eight items to measure the participant's mental health such as "I feel part of a group of friends" or "I worry about what other people will think of me even when I know it doesn't make any difference." The results from the experiment showed that groups from the "Virtual" section ended up experiencing higher levels of loneliness, social anxiety, and depression in the table (Cingel et. al., 2022). One thing to note is that females ended up on average having

high levels as opposed to males. There are possible limitations in the data such as only documenting a short portion of the pandemic spanning only a month long section that could use further analysis. Even though this model was used primarily to predict mental disorders, it has also been used to predict other outcomes that are related to issues with adolescents. Regardless, this study supports the idea that digital learning negatively affects adolescent mental health.

The impacts of digital learning were not only limited to the United States. In a paper by Debora Marques et al. 2020, a review was conducted over a variety of individual research papers in areas of China such as Shanghai, Hubei province, and other urban areas (Miranda et. al., 2020). In one of these research papers by Xinyan Xie et al., 2020, they surveyed over 2330 students through an online form inquiring about depressive and anxiety symptoms (Xie et. al., 2020). The responses were that 18.9% of students had anxiety symptoms and 22.6%

had depressive symptoms. As addressed in their limitations, these factors could have been caused by academic stress or other stressors. However, the impact the pandemic had on these adolescents was substantial as grades 2 to 6 all reported similar levels of stress.

Lack of Social Cues in Digital Learning

   The lack of social cues is also prevalent in digital learning. In a paper by John Orlando, a lack of social cues are mainly caused by the prevalence of text based communication in an online setting Orlando, 2017). He argues that even if text based communication may convey the same information, it is interpreted more harshly due to the fact that the communication lacks facial and verbal signals that we use to modulate the tone of our communication. This relates to the Social Learning Theory as stated previously, when applied to academic learning, most students would learn by example whether it be from the instructor or teacher in direct contact, not just from text. As a substitute, emoticons (emojis) are used as a weak recovery of the lost fidelity in moving from face to text. As stated above in the development of social skills section, humans rely on a variety of stimuli such as facial expressions, gestures, and body language which can not be replaced by emojis. When students are being deprived of these social skills, it can lead to social isolation which has a wide range of problems. Some of these problems include but are not limited to, feelings of loneliness, craving for social contact, and decreased happiness (Orben et. al., 2020). On a neurological level, isolation alters neural patterns in ways that are similar to other types of deprivation such as water or food deprivation (Tomova et. al., 2020). However some limitations to this is that the studies have been conducted on animals as opposed to humans.

Neural Development during Digital Learning

   The primary force behind neural development is synaptogenesis and neurogenesis. These are crucial in adolescent development because they are directly correlated with higher intelligence. In a research paper by Cai Qi et al., 2022 they discuss how the cognitive ability of humans are directly correlated to the increase in the number of synapses generated (Qi et. al., 2022). As previously mentioned, synaptogenesis and neurogenesis during adolescence is essential for development of the prefrontal cortex, and their resulting social capabilities. As such, social cue development in adolescents correlates to their intellectual development. As social cues are lacking during digital learning due to absence of a visual stimulus, the development of such skills would be attenuated. This causes the brain to have underdeveloped visual-social skills, which may result in a lack of synaptogenesis and neurogenesis in brain areas involved in these functions.

Mitigating Strategies

   With the negative effects of the Covid-19 pandemic being established, some strategies to mitigate future scenarios of digital learning would be to incorporate Social-Emotional Learning (SEL) into the learning (Weissberg et. al., 2015). Social and

emotional learning is tailored to the individual's strengths and weaknesses and adapts accordingly to them. This process starts as early as birth and then develops through the education of a student. Relationship skills are an integral part of SEL. This can be applied to a digital classroom by having students establish and maintain health relationships with one another. This includes being able to communicate clearly, cooperate, work collaboratively, listen actively, navigate settings with different social standards, and show leadership skills. In SEL, they have standards that apply these things. Even though SEL has been present for a while, it hasn't been applied to digital classrooms or hybrid learning and has only been used for in-person learning. Another possible solution would be to give adequate teacher training to prepare them for digital learning. Premade online curriculums that have built-in online interaction for peer to peer education would work as a healthy alternative that gives both the convenience and benefits of online learning.

Limitations

Possible limitations of these findings are that a narrow time frame was used to study the adolescents' response to digital learning. Researchers did not have enough time to fully understand the extent of which online learning impacted social skills in adolescents. The immediate effects are apparent, but long term studies are needed to assess the lasting effects that digital learning causes on social development. Future studies should confirm whether the detrimental impacts of digital learning on social skills persists beyond adolescence. With a possible confounding variable present in these studies, more
precise studies should be conducted in order to confirm these current studies. In one case, an online survey was conducted for a study due to the risk of spreading the virus (Xie et. al., 2020). The online element of this survey could have impacted how the students would have responded. In addition, these surveys restricted possible answer options for adolescents to show what they were experiencing by having them in a multiple-choice format.

Conclusion

The key insights provided by this paper revealed how adolescents during the earlier stages of the COVID-19 pandemic had their social skills impacted due to digital learning. With this review narrowing into the specific challenges adolescents faced, it presents the idea that adolescents were negatively impacted by digital learning through complex synaptic connections of the brain not activated for social skills, thus leading to a variety of mental issues. Although, the narrow time frame could affect the results of these studies. Potential measures for preventing a similar scenario would be by integrating Social-Emotional Learning into the digital classroom. This paper sheds light on the negative effects digital learning had on adolescents' social skills during COVID-19, and informs on measures to be taken by school administrators and teachers to subsidize the impact of digital learning on future generations.

Works Cited

1. New global tracker to measure pandemic's impact on education worldwide. https://www.worldbank.org/en/news/press-release/2021/03/26/new-global-tracker -to-measure-pandemic-s-impact-on-education-worldwide (2021).

2. Depression (major depressive disorder). https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/sy c-20356007 (2022).

3. Social Skills. https://www.mtu.edu/well-being/resource-tools/self-help-resources/mental-health resources/social-skills/ (2022).

4. CDC Museum COVID-19 Timeline. https://www.cdc.gov/museum/timeline/covid19.html (2023).

5. Definition of COVID-19. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/covid-19 (2020).

6. Coronavirus (COVID-19) roundup.https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocial care/conditionsanddiseases/articles/coronaviruscovid19roundup/latest

7. Murali Krishnan. India's digital divide grows among rural women. https://www.dw.com/en/indias-digital-divide-grows-among-rural-women/a-559490 74 (2020).

8. UNICEF. COVID-19: Schools for more than 168 million children globally have been completely closed for almost a full year, says UNICEF. https://www.unicef.org/india/press-releases/covid-19-schools-more-168-million-ch ildren-globally-have-been-completely-closed (2021).

9. Choosing an LMS: Canvas vs Moodle vs Blackboard vs LMS365. https://www.overtsoftware.com/canvas-vs-moodle-vs-blackboard-vs-lms365/ (2024).

10. What is Digital Learning? 10 Benefits of Digital Learning. https://lmshero.com/digital-learning/ (2024).

11. P. Soto-Icaza, F. Aboitiz, P. Billeke. Development of social skills in children: neural and behavioral evidence for the elaboration of cognitive models. Front. Neurosci. 9, 333 (2015).

12. Rehman, N. Mahabadi, T. Sanvictores, C. Rehman. Classical Conditioning. StatPearls [Internet]. (2024).

13. How Social Learning Theory Works. https://hr.berkeley.edu/grow/grow-your-community/wisdom-caf%C3%A9-wednes day/how-social-learning-theory-works (2024).

14. Orben, L. Tomova, S. J. Blakemore.The effects of social deprivation on adolescent development and mental health. The Lancet. Child & adolescent health. 4(8), 634–640 (2020).

15. B. Lam, S. M. McHale, A. C Crouter. Time with peers from middle childhood to late adolescence: developmental course and adjustment correlates. Child development. 85(4), 1677–1693 (2014).

16. J. Sakai. Core Concept: How synaptic pruning shapes neural wiring during development and, possibly, in disease. Proceedings of the National Academy of Sciences of the United States of America. 117(28), 16096–16099 (2020).

17. R. B. Adams, D. N. Jr, Albohn, K. Kveraga. Social Vision: Applying a Social-Functional Approach to Face and Expression Perception. Current directions in psychological science. 26(3), 243–248 (2017).

18. The Importance of Effective Communication. https://www.stevenson.edu/online/about-us/news/importance-effective-communication/

19. D. P. Cingel, A. R. Lauricella, L. B. Taylor, H. R. Stevens, S. M. Coyne, E. Wartella. U.S. adolescents' attitudes toward school, social connection, media use, and mental health during the COVID-19 pandemic: Differences as a function of gender identity and school context. PloS one. 17(10), e0276737 (2022).

20. D. Marques de Miranda, B. da Silva Athanasio, A. C. Sena Oliveira, A. C. Simoes-E-Silva. How is COVID-19 pandemic impacting mental health of children and adolescents?. International journal of disaster risk reduction : IJDRR, 51, 101845 (2020).

21. X. Xie, Q. Xue, Y. Zhou, K. Zhu, Q. Liu, J. Zhang, R. Song. Mental health status among children in home confinement during the Coronavirus disease 2019 outbreak in Hubei Province, China. JAMA Pediatr. 174(9), 898–900 (2020).

22. J. Orlando. Social cues in the online classroom. Online Classroom, 17, 10 (2017).

23. L. Tomova, K. L. Wang, T. Thompson, G. A. Matthews, A. Takahashi, K. M. Tye, R. Saxe. Acute social isolation evokes midbrain craving responses similar to hunger. Nature neuroscience, 23(12), 1597–1605 (2020).

24. C. Qi, L. D. Luo, I. Feng, S. Ma. Molecular mechanisms of synaptogenesis. Frontiers in synaptic neuroscience, 14, 939793 (2022).

25. R. P. Weissberg, J. A. Durlak, C. E. Domitrovich, T. P. Gullotta. Social and emotional learning: Past, present, and future. Handbook of Social and Emotional Learning: Research and Practice, 3–19 (2015).

Mechanisms of Resistance to Monoclonal Antibody Therapies in Cancer: Molecular, Cellular, and Tumor Microenvironment Factors Contributing to Therapeutic Evasion By Fia Inamdar

Abstract

Monoclonal antibody (mAb) therapies have emerged as a key component of cancer treatment by offering a targeted and precise approach to combating various types of malignancies. Unlike conventional treatments such as chemotherapy, mAbs are engineered to recognize and bind to specific proteins (antigens) expressed on cancer cells, allowing for minimal damage to healthy tissues. These therapies are widely used in cancers such as HER2-positive breast cancer, B-cell lymphomas, colorectal cancer, and non-small cell lung cancer (80). Through immune activation, inhibition of growth signaling, and delivery of cytotoxic agents, mAbs have significantly improved patient survival rates and quality of life.

While monoclonal antibody (mAb) therapies are effective, their long-term success is hindered by the development of resistance. Studies suggest that approximately 30-50% of patients treated with mAbs eventually develop resistance due to various mechanisms, limiting the therapy's efficacy (20). This complex resistance arises from multiple molecular, cellular, and environmental factors. These combined factors allow cancer cells to escape therapeutic targeting and continue to grow.

Recent findings highlight innovative strategies to overcome these barriers. Next-generation mAbs, including bispecific antibodies and antibody-drug conjugates (ADCs), are designed to address antigen heterogeneity and improve delivery (24). Advances in personalized medicine, guided by genomic and proteomic profiling, enable treatment customization based on the unique characteristics of individual tumors (33). This review aims to comprehensively understand the mechanisms driving resistance to mAb therapies and explore emerging solutions to mitigate these challenges. By bridging the gap between research findings and clinical applications, the findings highlight the importance of a multidisciplinary approach, which leverages molecular biology, immunology, and pharmacology to combat resistance and enhance the therapeutic potential of mAbs in oncology. Through addressing resistance, the field can move closer to achieving the full potential of monoclonal antibodies in the fight against cancer.

1Fia Inamdar, Kodaikanal International School, Joy Sapphire, N.S. Road No. 6, Vile Parle (W), Mumbai 400056, Maharashtra, India, fiainamdar@outlook.com

Table 1: Table of Contents

7.5 Targeting Emerging Resistance Mechanisms
7.6 Nanoparticle Delivery Systems
7.7 Next-Generation Monoclonal Antibodies: Leveraging Protein Dynamics to Overcome Resistance
7.8 Real-Time Adaptive mAb Therapies

1 Introduction

Cancer remains one of the most formidable challenges in modern medicine, accounting for millions of deaths worldwide each year. With over 2.04 million new cases reported in 2025, the need for effective and targeted cancer therapies is more urgent than ever. Among the transformative advancements in oncology, monoclonal mAb therapies have become a cornerstone of precision medicine, offering targeted solutions with remarkable efficacy and fewer off-target effects than traditional treatments, such as chemotherapy and radiation. These biologic agents, such as trastuzumab (Targeting HER2), cetuximab (targeting EGFR), and rituximab (Targeting CD20), have revolutionized the treatment landscape for various malignancies, improving patient outcomes and extending survival (71).

1.1 Monoclonal Antibodies, Production and Application

Monoclonal antibodies are designed to recognize and bind to a specific epitope on a target antigen, typically a protein or receptor overexpressed on cancer cells. This specificity underpins their utility in targeting antigens such as HER2 in HER2-positive breast cancer (84), EGFR in colorectal and lung cancers (8), and PD-L1 in melanoma and non-small cell lung cancer (13). The specificity of mAbs enables precise disruption of tumor cell signaling pathways, inhibition of angiogenesis, and activation of immune-mediated cytotoxicity, making them powerful tools in the oncology arsenal. Their ability to engage the immune system while sparing healthy tissues highlights their therapeutic advantage over conventional treatments (30).
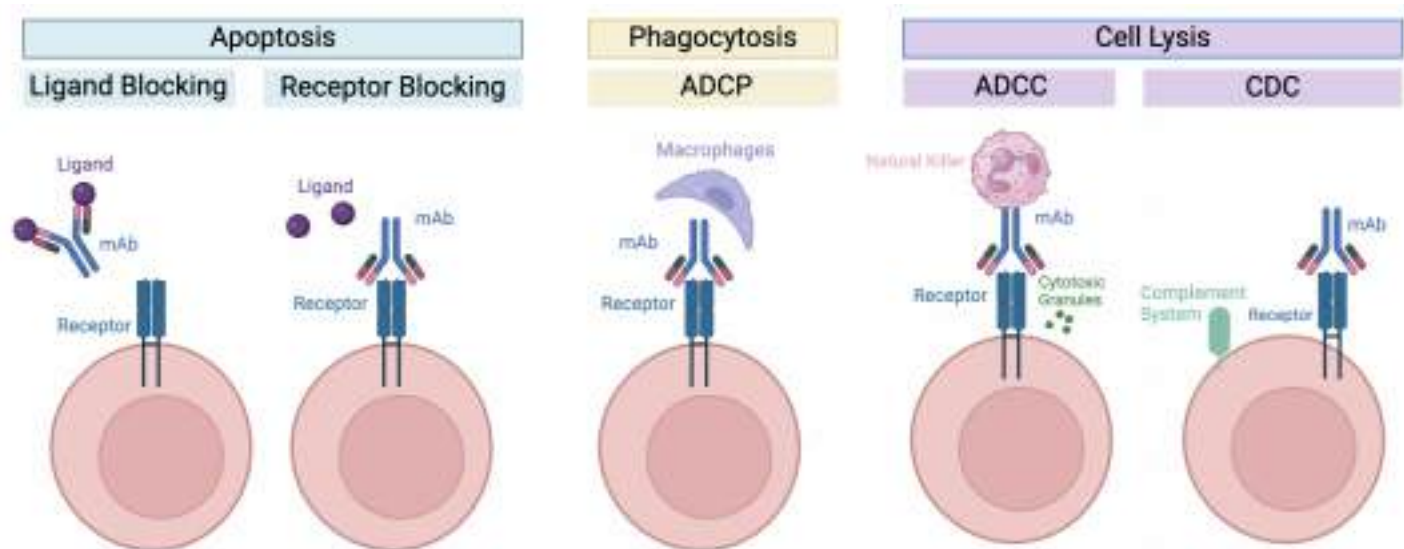
Figure 1: Mechanisms of Action of mAbs: Apoptosis, Phagocytosis, and Cell Lysis (30). This figure was created using BioRender. Apoptosis - mAbs can stop cancer cells from receiving survival signals by blocking either the ligands (small molecules that activate receptors) or the receptors on the cancer cells themselves. Without these signals, the cancer cells undergo programmed death. Phagocytosis (Antibody-Dependent Cellular Phagocytosis) - mAbs attach to the surface of cancer cells, making it easier for immune cells, such as macrophages, to recognize, engulf, and destroy the cancer cells. Cell Lysis- mAbs break apart cancer cells in two ways. First is ADCC (Antibody-Dependent Cellular Cytotoxicity), in which mAbs coat the cancer cells, allowing natural killer (NK) cells to bind to them. NK cells then release toxic substances, such as perforin, which creates holes in the cancer cell membrane, and granzymes, which enter the cell and trigger its death. Second, CDC (Complement-Dependent Cytotoxicity) under this mAb activates the complement system, a chain reaction of proteins in the immune system. This leads to the formation of the membrane attack complex (MAC), which punches holes in the cancer cell's membrane, causing it to break apart. (81)

This paper explores the **mechanisms of resistance** to **monoclonal antibody (mAb) therapies** in cancer, with a focus on the **molecular alterations** within cancer cells, such as **mutations in antigenic targets** (e.g., HER2 or EGFR) and **antigen downregulation**, which reduce the binding efficacy of therapeutic antibodies. It also examines **cellular adaptations**, including the emergence of **cancer stem cells (CSCs)**—a subset of tumor cells capable of self-renewal and inherent resistance to therapy—and the upregulation of **immune checkpoint proteins**, such as **PD-L1**, which allow cancer cells to evade immune detection (13). Additionally, the paper investigates the **tumor microenvironment (TME)**, a complex network of stromal cells, immune cells, and extracellular matrix components that supports tumor growth and creates physical and biochemical barriers to mAb penetration and efficacy. Key elements of the TME, such as **hypoxia** (low oxygen levels) and **immunosuppressive cells** like **regulatory T-cells (Tregs)** and **myeloid-derived suppressor cells (MDSCs)**, are discussed for their roles in therapeutic resistance.

By analyzing these factors, this study aims to gain a comprehensive understanding of the multifaceted challenges posed by resistance to mAb therapies (64, 13). It also proposes innovative strategies to overcome these barriers, including **bispecific antibodies**, which are engineered to target two different antigens simultaneously, thereby addressing antigen heterogeneity and improving therapeutic precision. Another approach involves **antibody-drug conjugates (ADCs)** (87, 11), which combine mAbs with potent cytotoxic

agents, delivering these directly to tumor cells to minimize damage to healthy tissues. Additionally, **combination therapies** are highlighted, where mAbs are paired with other treatments to enhance efficacy. For example, trastuzumab (anti-HER2) (84) is combined with pembrolizumab (anti-PD-1) (80) to target HER2-positive cancer cells while also reactivating the immune system's T cells. These integrative strategies show potential to overcome resistance mechanisms and enhance the durability and long-term success of mAb-based cancer treatments. Through this integrative approach, this review aims to provide a comprehensive analysis of mAb therapies, exploring their mechanisms of action, production processes, and the challenges posed by resistance. By integrating recent advancements in antibody engineering and therapeutic innovation, this study aims to address gaps in understanding and contribute to the development of more effective and durable cancer treatments.

2 Materials and Methods

This review paper was developed through a comprehensive literature analysis to explore the resistance mechanisms to mAb therapies in cancer. A systematic search was conducted across scientific databases, including PubMed, ScienceDirect, and Google Scholar, using key terms such as "mAbs," "therapeutic resistance," "tumor microenvironment," "immune checkpoints," and "HER2 resistance." Peer-reviewed articles, clinical trial reports, and review papers published between 2000 and 2023 were included to ensure the relevance and timeliness of the information.

The collected literature was critically evaluated to identify key molecular, cellular, and tumor microenvironmental resistance mechanisms. Data on innovative therapeutic strategies, including bispecific antibodies, antibody-drug conjugates (ADCs), and combination therapies, were extracted and synthesized to propose solutions for overcoming resistance.

The paper emphasizes evidence-based findings and incorporates data from clinical trials and preclinical studies to maintain objectivity and reliability. Comparative analysis was employed to identify patterns and gaps in existing research. Visual aids, such as tables and diagrams, were developed to summarize complex concepts. Integrating diverse perspectives ensures a holistic understanding of therapeutic resistance, allowing this paper to make a meaningful contribution to ongoing research in cancer treatment.

3 Background

mAbs are highly effective in cancer treatment due to their ability to employ direct and immune-mediated mechanisms to combat tumor growth and survival. One primary mechanism is the **direct targeting of cancer cells**, where mAbs bind specifically to receptors or antigens on the surface of tumor cells, disrupting critical pathways essential for their proliferation and survival. For instance, cetuximab, an anti-EGFR mAb, blocks the epidermal growth factor receptor (EGFR), thereby inhibiting downstream signaling through the RAS/RAF/MEK and PI3K/AKT pathways—pathways that are crucial for cell division and survival (62)

In addition to direct targeting, mAbs enhance the immune system's ability to eliminate cancer cells by activating the **immune system**. This occurs via **antibody-dependent cellular cytotoxicity (ADCC)** and **complement-dependent cytotoxicity (CDC)**. In ADCC, mAbs like rituximab, which target CD20 on B cells, recruit immune effector cells, such as natural killer (NK) cells and macrophages, to destroy the bound cancer cells (63). CDC involves activating the complement system, a part of the immune system that enhances the ability of antibodies to clear pathogens and damaged cells, further aiding in tumor destruction.

Another important mechanism is **checkpoint inhibition**, employed by a subclass of mAbs known as **immune checkpoint inhibitors**. These antibodies disrupt inhibitory signals that cancer cells use to escape detection by the immune system. For example, nivolumab, an anti-PD-1 antibody, and atezolizumab, an anti-PD-L1 antibody, block these inhibitory pathways, reactivating T-cells to recognize and attack cancer cells. By restoring T-cell activity, checkpoint inhibitors play a pivotal role in enhancing the body's natural anti-tumor immune responses (12).

Finally, mAbs can serve as delivery vehicles in the form of **antibody-drug conjugates (ADCs)**. ADCs are engineered by linking mAbs with potent cytotoxic agents, delivering these agents directly to cancer cells while sparing healthy tissues. A prime example is trastuzumab emtansine, which combines trastuzumab, an anti-HER2 antibody, with emtansine, a cytotoxic agent, to kill HER2-overexpressing cancer cells selectively (87). This targeted delivery minimizes systemic toxicity and enhances the precision of cancer therapy.
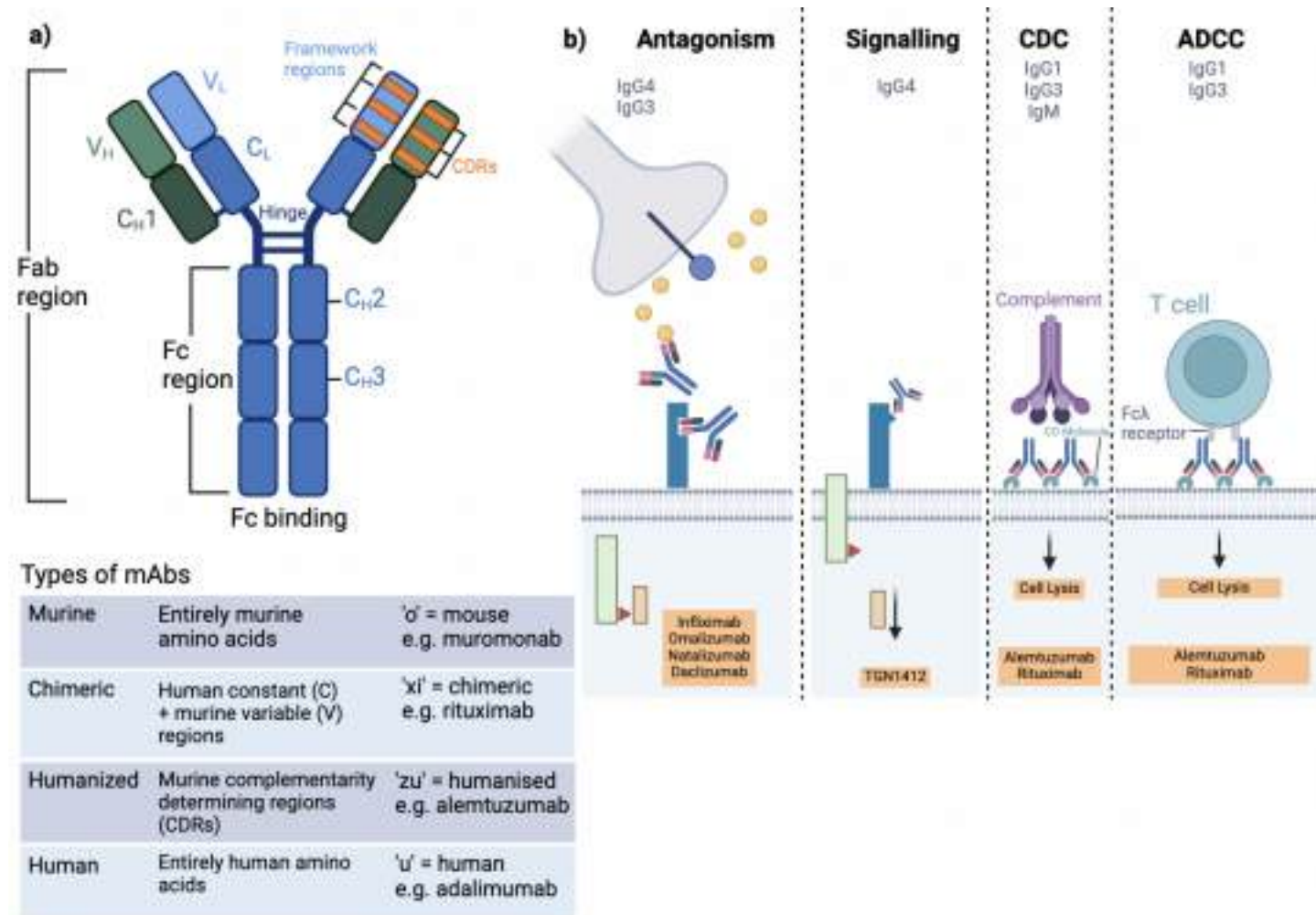

Figure 2: Development of mAbs: structure and function. Schematic structure of an immunoglobulin G (IgG) mAb (38). This figure was created using BioRender.

3.1 Mechanism of Resistance in mAb Therapies

The phenomenon of resistance to mAb therapies poses a significant challenge to their long-term efficacy in cancer treatment. Resistance mechanisms arise from intrinsic changes within tumor cells and extrinsic factors within the tumor microenvironment (TME). Understanding these mechanisms is crucial for improving the efficacy and durability of mAb-based treatments. Resistance, in this context, refers to the ability of cancer cells to evade or adapt to therapeutic agents, rendering the treatment less effective or ineffective over time (8). Studies have reported resistance in up to 30-50% of patients treated with trastuzumab for HER2-positive breast cancer and similar patterns in patients treated with rituximab for B-cell lymphomas (100). These cases highlight the urgent need to understand and address resistance to sustain the benefits of mAb therapies.
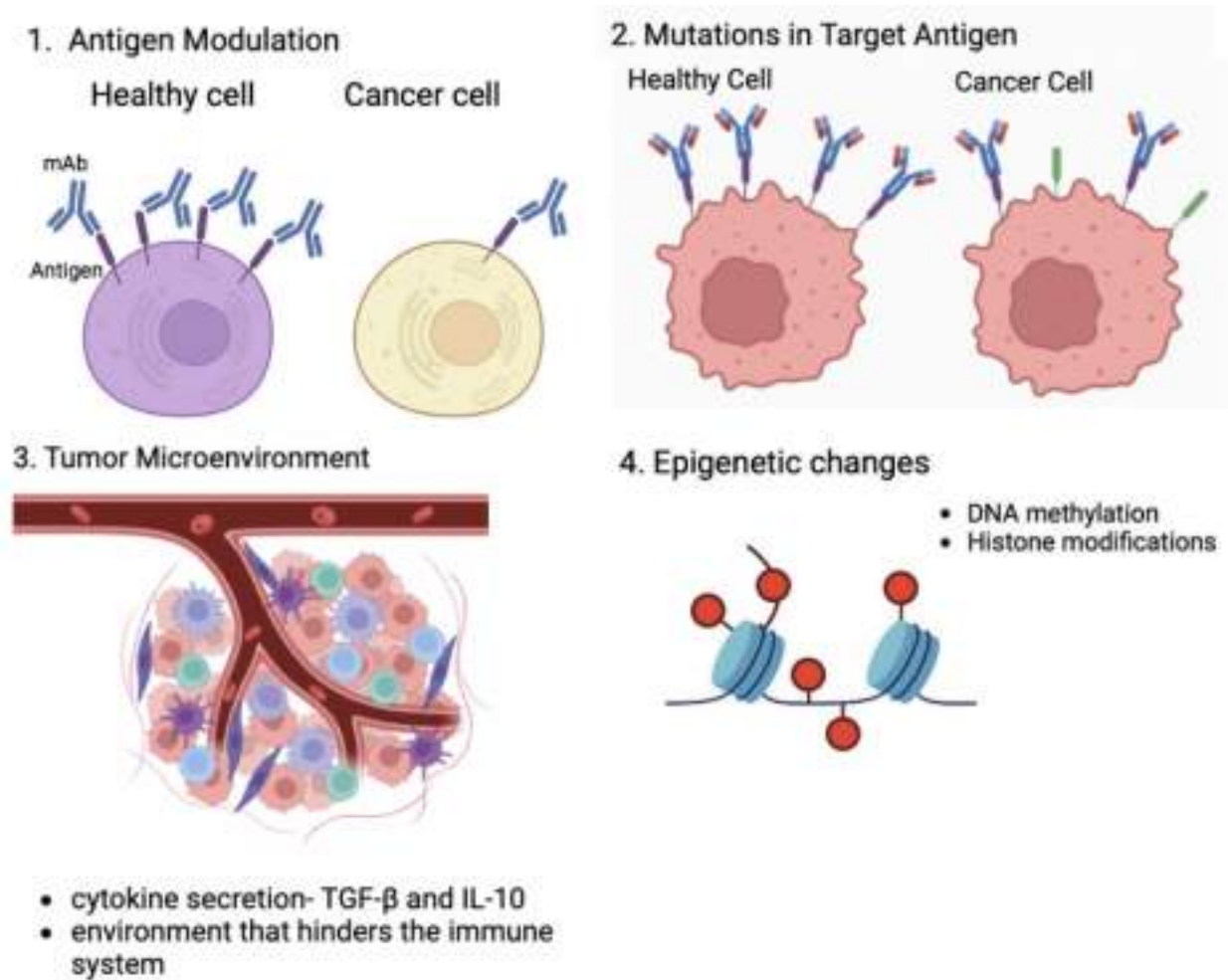

Figure 3: Mechanisms of resistance to mAb. This figure was created using BioRender. The main resistance mechanisms to mAb cancer therapies include antigen modulation, target antigen mutations, tumor microenvironment factors, and epigenetic changes, such as DNA methylation and histone modifications.

The **tumor microenvironment (TME)** further amplifies resistance. The TME consists of non-cancerous cells, such as stromal and immune cells, the extracellular matrix (ECM), and signaling molecules that surround and support the tumor (40). **Physical barriers**, such as a dense extracellular matrix (ECM), hinder the penetration of therapeutic antibodies into the tumor core. **Immunosuppressive cells,** such as regulatory T-cells (Tregs) and myeloid-derived suppressor cells (MDSCs), secrete factors that weaken the immune response, shielding the cancer cells. Additionally, **hypoxia**, or low oxygen levels within the tumor, promotes genetic instability and metabolic adaptations, enabling cancer cells to thrive despite therapy.

Table 2: Types of Resistance to Monoclonal Antibody Therapies in Cancer

Type of Resistance	Description and Mechanism	Examples	Result
Antigen Loss (3)	- Complete reduction or disappearance of the target antigen on the surface of cancer cells (47)	HER-2 (Human Epidermal Growth Factor Receptor 2) -loss in breast cancer (69)	- Cancer cells become unrecognisable to the immune system or mAb therapies because the
	- Downregulation, Genetic mutations of the target, epigenetic changes, tumor heterogeneity, or antigen shedding		target antigen is no longer present
Antigen Downregulation or Alteration (57)	- Loss or modification of the target antigen, preventing mAb binding.	HER2 downregulation in breast cancer, and EGFR mutations in lung cancer	- Reduced efficacy of targeted mAbs like trastuzumab (HER2) and cetuximab (EGFR).
Genetic / Epigenetic Modifications (77)	DNA methylation and histone modifications can alter the expression of genes involved in antigen presentation or immune responses.	Methylation of MGMT in glioblastoma	- Tumor cells evade immune recognition. Loss of antigen expression, resulting in reduced binding of mAbs.
Drug Efflux Mechanisms (105)	- Overexpression of efflux transporters that actively pump therapeutic antibodies out of cells.	P-glycoprotein (MDR1) in ovarian and breast cancer	- Decreased intracellular drug concentration, reducing mAb effectiveness.

Fc-Mediated Resistance/Fc Receptor Downregulation (7)	- Tumors alter Fc receptor expression to block ADCC, reducing immune system activation.	Low FcγR expression in lymphoma	- Impaired ADCC and reduced macrophage-mediated tumor killing. - Rituximab loses its immune-stimulating effect.
Activation of Alternative Pathways Or Bypass Pathways (68)	Cancer cells activate compensatory signaling pathways, which bypass the blockade by mAbs. MET, IGF-1R activation	PI3K/AKT/mTOR activation in breast cancer, MAPK pathway activation in colorectal cancer EGFR-mutant lung cancer	- mAb therapy fails as tumors use alternative survival mechanisms. - Leads to persistent tumor growth despite therapy.

Upregulation of - Tumor cells increase PD-L1 overexpression in - Prevents T-cell

Immune Checkpoints (Immune Evasion)/ Immune Checkpoint Activation	checkpoint molecules, and Upregulation of immune checkpoint proteins suppresses immune response, leading to mAb resistance. (79)	lung cancer and melanoma	activation and immune-mediated killing. - Leads to resistance against immune checkpoint inhibitors (anti-PD-1, anti-CTLA-4).
ADC Resistance	- Reduced internalization of ADCs or increased drug efflux by P-gp. (105)	T-DM1 resistance in HER2-positive breast cancer	- Reduces ADCC and overall immune-mediated tumor clearance. (14) - Failure to deliver cytotoxic payload to tumor cells. - Increased drug efflux pumps prevent intracellular

| Tumor Microenvironment (T ME)-Mediated | - The environment surrounding the tumor cells (tumor microenvironment) can become more resilient to therapies

A supportive environment actively contributes to the failure of mAb therapies.
Hypoxia, ECM barrier (40) | - Glioblastoma, pancreatic cancer
- The dense stromal environment in pancreatic tumors can create a physical barrier to drug delivery and provide survival signals to cancer cells | accumulation of the toxic payload.

- Reduced effectiveness of mAbs that rely on immune effector mechanisms like ADCC (antibody-dependent cellular cytotoxicity) and CDC (complement-dependent cytotoxicity). |

3.1.1 Tumor Cell-Intrinsic Mechanisms

(A) Antigen Loss or Alteration

The effectiveness of mAb therapies depends on the presence of specific antigens on tumor cells that serve as binding sites for antibodies. Resistance often arises due to the downregulation or complete loss of the target antigen. Tumor cells achieve this through genetic or epigenetic changes, rendering the mAb unable to bind effectively (92). For example, mutations in HER2 or EGFR alter the antigen's structure, reducing the binding affinity of mAbs like trastuzumab or cetuximab. These changes enable tumor cells to evade therapy and drive disease progression (47).

(B) Altered Signaling Pathways

Cancer treatment is further complicated by cancer stem cells (CSCs)—a rare subset of tumor cells with the unique ability to self-renew and resist conventional therapies. These cells often survive initial treatments, serving as a persistent reservoir for tumor regrowth and metastasis. Tumor cells also exploit immune regulatory pathways by upregulating immune checkpoint proteins such as PD-L1, which effectively suppress the immune system's ability to attack cancer cells (3). This immune evasion mechanism significantly reduces the efficacy of immune-modulating mAbs like nivolumab and pembrolizumab, which are designed to block PD-1/PD-L1 interactions and restore the immune system's anti-tumor activity.

In addition to immune suppression, tumor cells adapt to mAb therapies by activating alternative signaling pathways, a phenomenon known as cellular resistance. For example, when mAbs like cetuximab, which target the epidermal growth factor receptor (EGFR), block one signaling route, cancer cells often compensate by upregulating related pathways involving molecules such as HER3 or MET, ensuring continued growth and

survival. Furthermore, mutations in downstream signaling molecules, such as KRAS or PI3K, can lead to the constant activation of oncogenic cascades (93). These mutations bypass the inhibitory effects of mAb therapies, rendering them less effective. Collectively, these adaptations allow tumor cells to evade therapeutic targeting and maintain their malignant behavior (52).

3.1.2 Checkpoint Inhibition and Immune Evasion

Immune checkpoint inhibitors, a subclass of mAbs, have revolutionized cancer therapy by counteracting the inhibitory signals that cancer cells exploit to evade immune detection. These therapies target checkpoint pathways, such as PD-1 (programmed cell death protein 1) and PD-L1 (programmed death ligand 1). Nivolumab and atezolizumab, for example, block these pathways, effectively reactivating T cells to recognize and attack cancer cells (35). However, despite their groundbreaking potential, these therapies face significant challenges due to the adaptive strategies that cancer cells use to evade immune responses.

Cancer cells overexpress checkpoint molecules, such as PD-L1, as a defense mechanism to suppress T-cell activity. This overexpression is often induced by inflammatory signals in the tumor microenvironment, particularly interferon-gamma (IFN-γ), which paradoxically boosts PD-L1 expression. This creates an "immune escape" loop where the immune system's attempts to attack the tumor inadvertently lead to increased immune suppression (79). The effectiveness of checkpoint inhibitors can also be diminished by the presence of other inhibitory molecules, such as CTLA-4 (cytotoxic T-lymphocyte-associated protein 4), TIM-3 (T-cell immunoglobulin and mucin domain-containing 3), and LAG-3 (lymphocyte activation gene 3), which act as alternative pathways to block T-cell activation (14).

In clinical settings, the frequency of immune evasion mechanisms varies among patients and different types of cancer. For example, approximately 40-50% of patients with non-small cell lung cancer (NSCLC) exhibit high PD-L1 expression, which correlates with immune checkpoint resistance and poorer outcomes if not adequately targeted. In melanoma, up to 30-40% of patients develop resistance to PD-1/PD-L1 inhibitors within the first year of treatment. Moreover, a significant percentage of patients (ranging from 20-50% depending on the cancer type) experience either primary resistance, where the therapy fails to elicit an initial response, or acquired resistance, where tumors adapt over time to evade the immune system despite initial effectiveness (99).

The complexity of this immune suppression is further compounded by the tumor microenvironment, which actively promotes the recruitment of regulatory T cells (Tregs) and the expansion of myeloid-derived suppressor cells (MDSCs). These immune cells contribute to an immunosuppressive milieu by releasing cytokines such as IL-10 and TGF-β, which further inhibit T-cell functionality (41). This multifaceted suppression highlights the need for combination therapies that target multiple resistance pathways, such as pairing checkpoint inhibitors with agents that block other immunosuppressive signals or augmenting immune activation through cytokine therapy.

3.1.3 Reduced Effector Function

mAbs rely on immune-mediated mechanisms such as antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC) to destroy tumor cells. Resistance arises when tumor cells reduce Fc receptor expression on effector cells or impair complement activation, weakening ADCC and CDC. These often emerge due to the selective pressure exerted by mAb therapy. Tumors are genetically and phenotypically heterogeneous, and exposure to immune-mediated therapies can drive the survival of resistant subpopulations. For example, Chronic inflammation in the tumor microenvironment leads to epigenetic changes or mutations that suppress Fc receptor function or complement activation. Tumor cells may actively recruit immunosuppressive cells, such as regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs), which secrete factors that impair ADCC and CDC (97).

3.1.4 Mechanisms of Reduced Effector Function

Decreased Fc Receptor Expression: Tumor cells can reduce the expression of Fc receptors on effector cells, such as natural killer (NK) cells and macrophages. This decrease in Fc receptor expression impairs the ability of effector cells to bind to the Fc region of mAbs, thereby reducing antibody-dependent cellular cytotoxicity (ADCC). Impaired Complement Activation: Tumor cells can also inhibit complement activation, which is essential for complement-dependent cytotoxicity (CDC). The complement system is a group of proteins that work together to help eliminate bodily pathogens (94). Impaired complement activation reduces the ability of mAbs to induce CDC.

Reduced effector function is a common mechanism of resistance to mAbs, occurring in approximately 20-30% of patients with cancer (1). However, the exact frequency and prevalence of reduced effector function vary depending on the type of cancer, the specific mAb used, and the patient population being studied. In patients with chronic lymphocytic leukemia (CLL), reduced effector function has been reported to occur in approximately 40% of patients treated with the mAb rituximab. In patients with non-Hodgkin lymphoma (NHL), reduced effector function has been reported to occur in approximately 25% of patients treated with the mAb rituximab (96).

3.2 Tumor Microenvironment-Related Mechanisms

3.2.1 Immune Suppression

The TME fosters resistance by creating an immunosuppressive environment. Tumors recruit regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs), which inhibit cytotoxic immune cells like T cells and NK cells. Furthermore, tumors secrete immunosuppressive cytokines such as TGF-β and IL-10, suppressing anti-tumor immune responses (23). Alongside these factors, the upregulation of immune checkpoints like PD-L1 ensures tumor survival and proliferation despite immune attacks.

3.2.2 Physical Barriers

The dense extracellular matrix (ECM) and stromal components within the TME act as physical barriers, limiting the ability of mAbs to reach their target antigens. For example, fibroblasts produce collagen, creating a protective shield around tumor cells (40). Additionally, tumors often develop regions of hypoxia, where low oxygen levels reduce mAb efficacy and induce adaptive changes in tumor cells, such as metabolic reprogramming and epithelial-to-mesenchymal transition (EMT)(24). These changes increase tumor cell resistance to therapy.

3.3 Drug Resistance Mechanisms

3.3.1 Drug Efflux and Uptake Challenges

Tumor cells employ multidrug resistance (MDR) proteins, such as P-glycoprotein (P-gp), to expel therapeutic agents, including mAbs, from their intracellular environment. This reduces their intracellular concentrations and efficacy. Additionally, impaired drug uptake due to decreased functionality of transport proteins further limits mAb effectiveness (19). These mechanisms ensure tumor cells remain unaffected despite therapeutic interventions.

3.3.2 Antibody-Drug Conjugates (ADCs)

Antibody-drug conjugates (ADCs) are engineered to overcome traditional resistance mechanisms. ADCs couple mAbs with potent cytotoxic agents, allowing targeted delivery of drugs directly to tumor cells while sparing normal tissues. For instance, trastuzumab emtansine (T-DM1) combines trastuzumab, an anti-HER2 mAb, with the cytotoxic agent emtansine. This dual-action approach enables ADCs to bypass challenges such as antigen loss by delivering their toxic payload regardless of reduced mAb efficacy (105). ADCs also help overcome physical barriers in the TME by precisely targeting cancer cells and ensuring drug delivery even in resistant environments (15).

4 Discussion

Specific resistance mechanisms

4.1 Altered Receptor Signaling: Overexpression, Phosphorylation, and Variant Forms

A major mechanism of resistance to monoclonal antibody (mAb) therapies involves alterations in receptor tyrosine kinases (RTKs), which are key regulators of cell growth, survival, and differentiation. These alterations may occur via overexpression and aberrant phosphorylation of alternative RTKs or through the production of structurally altered receptor variants.

In some tumors, alternative RTKs such as MET (hepatocyte growth factor receptor), IGF-1R (insulin-like growth factor receptor 1), and ERBB3 (a member of the HER/EGFR family) are overexpressed. These receptors compensate for the blockade of primary targets by continuing to activate downstream pathways such as PI3K/AKT and MAPK, which support proliferation and inhibit apoptosis (60). For instance, when cetuximab targets EGFR, tumor cells may activate MET or ERBB3 to bypass the inhibited pathway. Aberrant phosphorylation—addition of phosphate groups in the absence of normal activating signals—can further stabilize these alternative receptors in their active conformations, contributing to sustained oncogenic signaling (70).

In parallel, cancer cells may produce receptor variants that are inherently resistant to mAb binding. A well-characterized example is EGFRvIII, a truncated mutant of EGFR that lacks the extracellular ligand-binding domain. EGFRvIII is constitutively active and evades inhibition by cetuximab, as the therapeutic antibody can no longer bind its target (45). This variant is commonly found in glioblastoma and other aggressive cancers and enhances downstream signaling even in the absence of ligand stimulation. These mutant forms not only evade therapeutic binding but also amplify oncogenic signaling, promoting tumor aggressiveness and treatment failure (4).

Given the role of both overexpressed alternative receptors and structurally modified variants in driving resistance, there is a need to develop therapeutics that can overcome these adaptations. Promising approaches include bispecific antibodies that simultaneously target the canonical receptor and its variant or alternative, as well as small-molecule inhibitors that bind intracellular kinase domains independent of extracellular alterations.

4.2 Increased Expression of the Target Receptor

Another common resistance mechanism is the overexpression of target receptors such as EGFR or HER2. These receptors are usually regulated by degradation pathways that remove excess proteins from the cell surface. However, in resistant cancer cells, these degradation pathways can become defective, leading to an accumulation of receptors on the cell surface (64). This overexpression saturates mAb therapies like trastuzumab (HER2-targeting) or cetuximab (EGFR-targeting), reducing effectiveness.

For instance, in HER2-positive breast cancer, the overexpression of HER2 may exceed the binding capacity of trastuzumab, allowing unbound receptors to continue driving cancer cell proliferation (84). Similarly, increased EGFR expression due to defective degradation can diminish cetuximab's ability to inhibit signaling pathways effectively.

4.3 Impact on Therapeutic Outcomes

Protein modifications and receptor variants complicate cancer treatment by creating redundancies in signaling pathways and evading targeted therapies. These mechanisms reduce the efficacy of existing mAbs and contribute to disease progression and poorer patient outcomes (86). Cancer cells' adaptability underscores the

importance of developing novel strategies, such as combination therapies that target multiple RTKs simultaneously or next-generation mAbs designed to recognize and neutralize mutant variants.

4.4 Alternative Pathway Activation

Cancer cells often activate alternative signaling pathways to bypass the blocked pathways targeted by mAb therapies. These bypass mechanisms ensure tumor cells' continued proliferation, survival, and resistance. One prominent example is the activation of the Src kinase pathway. Src, a non-receptor tyrosine kinase, can bypass the epidermal growth factor receptor (EGFR) inhibition by activating downstream signaling cascades, including the PI3K/AKT and MAPK/ERK pathways. These cascades promote cellular proliferation and survival, thereby diminishing the effectiveness of EGFR-targeted therapies such as cetuximab and panitumumab (111).

Similarly, tumors overexpressing transforming growth factor-alpha (TGFα) can maintain EGFR activity independent of ligand binding, contributing to resistance against EGFR-targeted therapies. This overexpression bypasses the blockade of EGFR but also amplifies tumor growth and survival signals.

Cancer cells often overexpress alternative pro-angiogenic factors in anti-angiogenic therapies targeting vascular endothelial growth factor (VEGF), such as bevacizumab (5). Fibroblast growth factor (FGF) and placental growth factor (PGF) are commonly upregulated in resistant tumors, sustaining angiogenesis despite VEGF inhibition. These factors activate alternative receptors, such as FGFR and VEGFR-1, ensuring the cancer receives an adequate blood supply for growth and survival (Fischer et al., 2010). This redundancy in angiogenic signaling pathways is particularly challenging, as it allows tumors to adapt dynamically to therapy (56).

4.5 Steric Hindrance by Other Cell Surface Proteins

Steric hindrance is a significant physical resistance mechanism wherein other cell surface proteins mask the target receptors, preventing the effective binding of mAbs. For instance, the protein MUC4, a membrane-associated mucin, is known to obscure the HER2 receptor on the surface of cancer cells. HER2 is a critical target in HER2-positive breast cancers, and its masking by MUC4 significantly reduces the ability of trastuzumab to bind effectively to its target. MUC4 physically hinders trastuzumab binding and interferes with receptor internalization and degradation, thereby sustaining HER2-mediated signaling. This resistance mechanism highlights the role of cell surface glycoproteins in modifying the accessibility of therapeutic targets. Studies have shown that silencing MUC4 expression can restore trastuzumab sensitivity in resistant cancer cells, emphasizing the potential of targeting such masking proteins to enhance mAb efficacy (37).

Additionally, steric hindrance is not limited to HER2-targeting therapies. Other examples include masking EGFR by overexpressed mucins or extracellular matrix proteins, which create a physical barrier that impedes antibody binding. The dense stromal environment often present in solid tumors can exacerbate these barriers, reducing therapeutic effectiveness (43).

4.6 Secretion of Alternative Ligands

Resistance to mAbs can also arise through the secretion of alternative ligands that bypass the inhibition of targeted pathways. In tumors treated with anti-angiogenic therapies like bevacizumab, cancer cells often increase the secretion of essential fibroblast growth factor (FGF), hepatocyte growth factor (HGF), and platelet-derived growth factor (PDGF). These angiogenic factors stimulate alternative signaling pathways that promote the formation of new blood vessels, ensuring that the tumor receives nutrients and oxygen despite VEGF blockade (36).

This mechanism is particularly problematic in tumors with high signaling plasticity, where multiple angiogenic factors can redundantly drive vascular growth. For example, HGF activates the MET receptor, which promotes angiogenesis and enhances tumor cell motility and invasion, further complicating treatment outcomes.

In addition to pro-angiogenic ligands, tumors may secrete ligands that activate alternative signaling receptors in the same family as the target receptor. For instance, in EGFR-resistant tumors, overexpression of ligands such as amphiregulin or epiregulin can activate other HER family receptors, such as HER3, bypassing EGFR inhibition and maintaining downstream signaling pathways essential for tumor survival.

4.7 Drug Influx/Efflux Mechanisms

One of the critical challenges in cancer therapy is the efficient delivery of therapeutic agents to tumor cells. Resistance mechanisms associated with drug influx (entry) and efflux (exit) have emerged as significant barriers, particularly for therapies involving mAbs and small-molecule inhibitors. These mechanisms reduce
the intracellular concentration of therapeutic agents, compromising their efficacy and allowing cancer cells to survive and proliferate (20). Two key components of these resistance mechanisms are multidrug resistance proteins (MDR) and the structural challenges of irregular tumor vasculature.

4.7.1 Multidrug Resistance Proteins (MDR)

MDR proteins are specialized transmembrane pumps that actively transport drugs out of cells, often against concentration gradients. These proteins belong to the ATP-binding cassette (ABC) transporter family, which uses the energy from ATP hydrolysis to drive the efflux of various substrates, including therapeutic agents. A prominent example is P-glycoprotein (P-gp)(106), also known as ABCB1, which has been extensively studied for its role in mediating drug resistance.

P-gp and other MDR proteins recognize and expel various chemotherapeutic drugs, small-molecule inhibitors, and even some mAbs, reducing their intracellular accumulation and effectiveness. For example, P-gp has been implicated in resistance to tyrosine kinase inhibitors (TKIs) like gefitinib and lapatinib, which are used to target EGFR and HER2, respectively. By expelling these inhibitors, P-gp decreases its ability to inhibit its targets effectively, allowing cancer cells to bypass therapeutic blockades.

The activity of MDR proteins is particularly challenging in tumors with high genomic instability, as these tumors often overexpress efflux pumps. This overexpression can be intrinsic (present before treatment) or acquired (developed during therapy). For instance, overexpression of P-gp in resistant cancer cells has been observed to correlate with a decrease in intracellular concentrations of drugs like imatinib, a TKI used to treat chronic myeloid leukemia (CML)(21). Studies have shown that blocking P-gp activity using inhibitors or RNA interference can restore drug sensitivity, highlighting the potential of targeting efflux pumps to overcome resistance.

4.7.2 Irregular Tumor Vasculature

The irregular structure and poor permeability of tumor vasculature pose significant challenges for drug delivery, particularly for large molecules like mAbs. Unlike normal blood vessels, tumor vessels are often disorganized, leaky, and poorly perfused, leading to uneven drug distribution within the tumor microenvironment. This phenomenon, the enhanced permeability and retention (EPR) effect, can paradoxically result in inadequate drug penetration in some tumor regions.

Tumor vasculature also exhibits high interstitial fluid pressure (IFP) due to the lack of proper lymphatic drainage. This elevated pressure creates a barrier to the passive diffusion of drugs from blood vessels into tumor tissues. For mAbs, which rely on systemic circulation to reach their targets, poor vascularization and high IFP significantly reduce their ability to penetrate deep into the tumor core, leaving regions of the tumor untreated. These untreated areas, often referred to as hypoxic zones, can foster the survival of therapy-resistant cancer cells and promote tumor relapse.

4.7.3 Combined Impact and Therapeutic Implications

The interplay between MDR proteins and irregular tumor vasculature creates a formidable barrier to effective drug delivery. Efflux pumps actively expel therapeutic agents from cells, while poor vascularization limits their ability even to reach the target cells in the first place. Together, these mechanisms reduce drug efficacy and create selective pressure that favors the survival of resistant cancer cell populations. To address these challenges, several strategies are being explored:

Efflux Pump Inhibitors: To restore intracellular drug concentrations, compounds that inhibit MDR proteins, such as verapamil and tariquidar, are being investigated. These inhibitors block the activity of efflux pumps like P-gp, allowing therapeutic agents to accumulate within cancer cells.

Nanoparticle-Based Drug Delivery: Nanoparticles and liposomes can encapsulate therapeutic agents, protecting them from efflux pumps and enhancing their penetration into tumors. For example, liposomal formulations of doxorubicin have shown improved distribution in tumors with irregular vasculature.

Vascular Normalization: Strategies to normalize tumor vasculature, such as using anti-angiogenic agents, aim to improve blood flow and reduce IFP. By creating a more organized vascular network, these approaches

enhance the delivery and distribution of therapeutic agents.

Combination Therapies: Combining MDR inhibitors with drugs that target the tumor microenvironment, such as anti-hypoxia agents, can simultaneously address efflux-mediated resistance and poor drug delivery (46).

4.8 Genomic Alterations and Mutations

Genomic alterations and mutations are a cornerstone of resistance mechanisms to mAb therapies. They enable cancer cells to evade targeted treatments. These mutations can be broadly categorized into **pre-existing** and **emergent**, each playing a distinct role in therapeutic resistance.

Pre-existing Mutations- Pre-existing mutations, or primary resistance mechanisms, are inherent genetic changes within a subpopulation of tumor cells present before treatment. These mutations confer a survival advantage under the selective pressure of mAb therapies, allowing resistant cancer cells to proliferate while sensitive cells are eliminated. For instance, mutations in the KRAS gene—a critical mediator of downstream signaling from the epidermal growth factor receptor (EGFR)—have been linked to resistance against EGFR-targeting antibodies such as cetuximab and panitumumab in colorectal cancer. Similarly, mutations in PIK3CA, which encodes a subunit of the PI3K enzyme, or BRAF, a serine/threonine kinase, disrupt the efficacy of targeted therapies by constitutively activating growth and survival pathways even in the presence of inhibitors (107).

Emergent Mutations- Emergent or secondary resistance mutations arise after therapy initiation, typically due to the selective pressure exerted by the treatment. These mutations do not abolish the
receptor's functionality but instead modify the structure of the drug-binding domain, reducing the binding affinity of mAbs while preserving the receptor's ability to signal. A classic example is the EGFR T790M mutation, observed in non-small cell lung cancer (NSCLC) patients treated with EGFR inhibitors. This mutation increases the receptor's affinity for ATP, reduces interaction with inhibitors, and allows signaling to continue unabated. Similarly, in chronic myeloid leukemia (CML), secondary mutations in the BCR-ABL fusion gene, such as the T315I mutation, alter the kinase domain to evade inhibition by drugs like imatinib (58). These mutations highlight the dynamic nature of cancer resistance, requiring constant innovation in therapeutic design.

4.9 Gene Amplification and Overexpression

Gene amplification and overexpression of critical oncogenes and receptors represent another pivotal mechanism by which tumors develop resistance to mAb therapies. Amplification refers to the increase in gene copies within the genome, leading to the overproduction of the corresponding protein.

Gene Amplification of RTKs- One of the most well-characterized examples of resistance through gene amplification involves receptor tyrosine kinases (RTKs) like MET and EGFR. In EGFR-mutant

NSCLC, tumors amplify the MET gene to compensate for EGFR inhibition. MET amplification activates downstream pathways such as the PI3K/AKT and MAPK signaling cascades, bypassing the blockade of EGFR by mAbs or tyrosine kinase inhibitors (TKIs). This phenomenon has been observed in patients who initially respond to EGFR-targeting therapies but later relapse due to the emergence of MET-driven signaling (66). Amplified EGFR itself can also overwhelm mAb therapies, as higher receptor densities on the tumor cell surface reduce the efficacy of receptor blockade.

Amplification of Downstream Effectors- In some cases, tumors amplify genes encoding downstream signaling molecules to sustain survival and proliferation despite upstream inhibition. For instance, amplifying PIK3CA or AKT1, both integral components of' pathways, allows component tumors to bypass receptor inhibition and continue transmitting growth signals. Similarly, amplification of KRAS enables persistent activation of the MAPK pathway, diminishing the impact of upstream therapies targeting EGFR or HER2 (58).

Clinical Implications and Strategies The amplification of these genes contributes to resistance and complicates treatment strategies. It often necessitates combination therapies, such as dual inhibition of the amplified pathway and its compensatory signaling routes. Targeting MET amplification, for example, has shown promise with MET inhibitors used in conjunction with EGFR-targeting agents in NSCLC. Additionally, emerging approaches like genomic profiling enable the identification of amplified genes, allowing personalized treatments that preemptively address resistance mechanisms.

4.10 KEGG Pathways with Resistance Mechanisms

respond to.

Recent insights from KEGG pathway analyses and integrated transcriptomic datasets have enabled the classification of tumors by resistance phenotypes, suggesting tailored mAb and immunotherapy strategies. The following table summarizes key resistance traits, associated tumor types, and which mAb strategies they are likely to respond to or not

Table 2: Tumor Resistance Characteristics and Therapeutic mAb Suitability

Tumor resistance characteristics	Tumors
Large number of mutations	melanoma, lung, colon
Low number of mutations	prostate, thyroid, glioblastoma?
Downregulate TAA, 'cold tumors'	HNSCC, lung, colorectal, bladder, laryngeal, breast
Downregulate P-gp expression	Breast, colorectal, prostate, schwannoma

OXPHOS defects	AML, pancreatic, prostate, glioblastoma 'cold tumors'	Altered amino acid metabolism	Breast, Kidney
	Near universal	Specific target antigens	Melanoma, Neuroblastoma, Colon, Lymphoma
Tumor heterogeneity, dense surrounding tissue, immunosuppressive microenvironment	Solid tumors	Heterogeneous tumors and 'hot' tumors	ALL, follicular lymphoma, Multiple myeloma
Altered lipid metabolism	Stomach		

> 2-fold DEG upregulated	Kidney, Pancreatic	Antibody-oligonucleotide conjugate (AOC) or Antibody-miRNA conjugate	
Dense surrounding tissue	NSCLC, Melanoma, Solid tumors	Radioimmunotherapy (Antibody-radioactive-drug-conjugates) due to abscopal effect	

Downregulate Antibody dependent cellular toxicity (ADCC)	Not yet investigated	Glycoengineered antibodies (eg. afucosylated) First gen mAB	

This table synthesizes tumor resistance traits with associated KEGG pathway data and therapeutic implications. For instance, tumors with high mutational burden like melanoma and NSCLC activate robust neoantigen presentation, favoring checkpoint inhibitors and first-line mAbs (29). In contrast, glioblastomas and prostate cancers with lower mutation rates may resist standard mAb-based immunotherapy due to limited antigenicity.

Downregulation of tumor-associated antigens (TAAs) or antigen processing components in "cold" tumors like

bladder or breast cancer limits response to standard mAbs. Multispecific antibodies or BiTEs may overcome this by bridging tumor antigens with effector T cells (112).

OXPHOS defects, a nearly universal feature in solid tumors, rewire metabolism toward glycolysis. In such cases, targeting glucose transporters like GLUT1 with antibody-drug strategies has shown promise (22). Similarly, tumors with dense stroma or high interstitial pressure benefit more from radioimmunotherapy or ADCs due to their superior tissue penetration and localized cytotoxicity.

KEGG enrichment of amino acid metabolism pathways (e.g., glutamine, leucine) in breast and renal cancers aligns with overexpression of transporters like SLC1A5 and SLC6A14, now explored as antibody targets (51).

Lastly, emerging work in glycoengineering (e.g., afucosylated antibodies) offers solutions for tumors that suppress ADCC. These engineered antibodies show enhanced binding to Fcγ receptors on NK cells (73).

5 Addressing the Impact of Resistance and Innovations in mAb Therapy

The impact of resistance to mAb therapies is profound, spanning clinical, economic, and scientific dimensions. Clinically, resistance results in disease progression, reduced survival rates, and limited patient treatment options. Economically, it escalates healthcare costs through the necessity for additional therapies, frequent diagnostic tests, and prolonged hospitalizations. Scientifically, it highlights the limitations of existing treatments and emphasizes the urgent need for continuous innovation in therapeutic development. Addressing these resistance mechanisms is critical for enhancing the durability and efficacy of mAb therapies and improving patient outcomes. Advanced antibody engineering offers promising solutions to these challenges, including bispecific antibodies such as blinatumomab and next-generation antibody-drug conjugates (ADCs) (67). Additionally, personalized medicine approaches, underpinned by genomic and proteomic profiling, pave the way for tailored treatment plans that address individual tumor characteristics.

Innovations in mAb design are central to overcoming resistance. Next-generation mAbs, such as bispecific antibodies, are engineered to simultaneously target multiple antigens, effectively addressing antigen heterogeneity and loss issues. For example, trastuzumab deruxtecan, an advanced ADC, combines the targeting precision of trastuzumab with a potent cytotoxic agent to bypass resistance mechanisms caused by HER2 mutations. ADCs also circumvent physical barriers within the tumor microenvironment (TME) by directly delivering cytotoxic payloads to cancer cells, irrespective of antigen density or accessibility. Personalized medicine further enhances these innovations by tailoring treatments to specific molecular and cellular resistance mechanisms identified through genomic sequencing and proteomic profiling (11). In HER2-positive breast cancer, for instance, identifying mutations conferring resistance enables the strategic use of therapies like trastuzumab emtansine or bispecific antibodies to target resistant tumors effectively.

These advancements underscore the necessity of integrating molecular insights into clinical practice.

Researchers and clinicians can refine therapeutic strategies by addressing key resistance mechanisms, such as the overexpression of alternative receptors, genomic alterations, immune evasion, and drug delivery challenges.

6 Limitations in mAb Research and Review

Despite their transformative impact in oncology, mAb therapies face significant limitations, particularly the development of resistance, which undermines their long-term efficacy. While mAbs offer advantages such as high specificity, immune system engagement, and reduced off-target effects compared to conventional therapies, approximately 30–50% of patients eventually develop resistance through mechanisms like antigen loss, signaling pathway bypass, or immune evasion (64, 13). Additionally, challenges such as high production costs, variable patient responses, and tumor microenvironment adaptations further limit their clinical utility (8, 11). A critical analysis of these limitations, supported by recent evidence, is essential to advance more durable and accessible mAb-based treatments.

6.1 Overemphasis on Clinical Successes and Underrepresentation of Real-World Data

Clinical trials for mAbs often showcase high efficacy rates under controlled conditions. For instance, trastuzumab has demonstrated significant survival benefits in HER2-positive breast cancer patients, with a 37% reduction in mortality when combined with chemotherapy (42). However, real-world data suggest that patients with advanced-stage disease, comorbidities, or treatment histories often experience diminished benefits. A study analyzing outcomes in community oncology practices revealed that up to 20% of HER2-positive patients did not respond to trastuzumab due to disease heterogeneity and acquired resistance (49). These discrepancies highlight the need for broader studies that reflect diverse patient populations.

6.2 Fragmented Understanding of Resistance Mechanisms

Resistance to mAb therapies arises from complex and overlapping mechanisms. Antigen loss, for example, was identified in colorectal cancer patients treated with cetuximab, where mutations in the EGFR extracellular domain (e.g., S492R mutation) rendered the therapy ineffective (17). Similarly, tumors adapt by activating compensatory pathways. MET amplification in EGFR-mutated non-small cell lung cancer was shown to bypass EGFR inhibition by erlotinib, a phenomenon likely relevant to mAb resistance. However, most studies explore these resistance mechanisms in isolation. A lack of integrative approaches fails to address the dynamic and multifactorial nature of resistance observed in clinical settings.

6.3 Limitations of Preclinical Models

While preclinical studies using animal models provide foundational insights, their translational value remains limited. For instance, genetically engineered mouse models used to evaluate PD-1/PD-L1 inhibitors failed to replicate the complex tumor microenvironment (TME) of human cancers. The failure of agents like bevacizumab to show survival benefits in ovarian cancer, despite promising preclinical results, underscores this limitation (26). Furthermore, ethical and logistical challenges prevent longitudinal human studies that could reveal chronic resistance mechanisms or cumulative toxicity.

6.4 Economic and Logistical Barriers

The high cost of mAb development limits the exploration of diverse therapeutic targets (72). Commercial interests further skew research priorities, as high-value targets like HER2 and PD-L1 receive disproportionate attention, leaving rare cancers or less lucrative targets underexplored. Additionally, negative trial results, such as the failure of ipilimumab in glioblastoma (27), often go unpublished, contributing to publication bias and an incomplete understanding of mAb limitations.

6.5 Inconsistencies in Data Reporting and Analysis

Systematic reviews and meta-analyses often struggle with heterogeneity in study designs. For example, response rates, progression-free survival (PFS), and overall survival (OS) metrics vary significantly across studies evaluating rituximab in lymphoma due to differences in patient populations (e.g., age, disease stage), treatment protocols (e.g., dosing schedules, combination therapies), and endpoint definitions, making robust comparisons challenging (18). Furthermore, the proprietary nature of many mAb platforms restricts access to raw data, hindering independent verification and broader collaboration.

6.6 Emerging Technologies with Untested Challenges

Next-generation mAbs, such as bispecific antibodies and antibody-drug conjugates (ADCs), offer innovative solutions but face significant challenges. Bispecific antibodies like blinatumomab (targeting CD19 and CD3) have shown promise in acute lymphoblastic leukemia but exhibit high rates of neurotoxicity and immune-related adverse effects (103). Similarly, ADCs like trastuzumab deruxtecan address HER2 resistance but present risks of interstitial lung disease. These issues highlight the need for more rigorous preclinical and clinical evaluations to balance innovation with safety.

6.7 Recommendations for Addressing Limitations

A multifaceted approach is essential to bridging these gaps. Expanding patient representation in clinical trials can enhance the generalizability of results, as demonstrated by efforts to include racially and ethnically diverse populations in PD-L1 inhibitor studies (50). Integrative research on resistance mechanisms, combining genomic, proteomic, and metabolomic profiling, can provide a more comprehensive understanding. Open-access initiatives, such as the Cancer Moonshot Data Initiative, facilitate collaboration and transparency and address the data-sharing challenges that hinder progress.

Table 3: Comparative table of resistance mechanisms across different cancer types

Cancer Type	Common Resistance Mechanisms	Examples of Resistance	Impact on Therapy	Potential Solutions

HER2-Positive Breast Cancer	- Antigen loss/downregulation (100, 64) - Activation of bypass pathways (PI3K/AKT/mTOR) (64, 5) - Immune evasion (upregulation of PD-L1) (13, 14)	- HER2 downregulation leading to trastuzumab resistance - PI3K/AKT mutations allow escape from HER2 inhibition	- Reduced efficacy of trastuzumab, pertuzumab, and T-DM1 - Tumor evades immune response	- Bispecific antibodies (targeting HER2 and PD-1) - PI3K inhibitors (e.g., alpelisib) (23, 103) - Antibody-drug conjugates (ADCs) (e.g., trastuzumab deruxtecan) (87, 11) -
Non-Small Cell Lung Cancer (NSCLC)	- EGFR mutations & bypass signaling (MET, ALK, KRAS activation) (8, 66) - Upregulation of immune checkpoints (PD-L1 overexpression) (13, 80)	- EGFR T790M mutation leads to cetuximab resistance - MET amplification bypasses EGFR inhibition	- Resistance to anti-EGFR therapies (cetuximab, osimertinib) - Immune checkpoint inhibitors (ICIs) fail due to immune evasion	Third-generation TKIs (osimertinib for T790M mutation) - Bispecific T-cell engaging antibodies (EGFR/MET dual inhibitors)
Colorectal Cancer (CRC)	- EGFR downregulation & RAS mutations (8) - Alternative pathways (MAPK, Wnt/β-catenin activation) (6)	- KRAS/NRAS mutations cause resistance to cetuximab and panitumumab - Wnt signaling promotes immune evasion	- Anti-EGFR therapy (cetuximab, panitumumab) becomes ineffective	- KRAS inhibitors (sotorasib for G12C mutation) - MEK inhibitors (for MAPK pathway activation) - Immune checkpoint inhibitors for MSI-H tumors

Hematologic Cancers (B-cell lymphomas, Leukemias)	- Antigen escape (CD20 downregulation in lymphoma) (78) - Overexpression of drug efflux pumps (P-gp in leukemias) (15, 106)	- Rituximab resistance in CD20-negative B-cell lymphoma - P-gp overexpression leading to ADC resistance (brentuximab vedotin)	- Loss of mAb binding targets - Chemoresistance due to drug efflux	- CD19/CD20 bispecific antibodies (e.g., blinatumomab) - Combining ADCs with efflux pump inhibitors
Glioblastoma (Brain Tumors)	- Tumor microenvironment (hypoxia, ECM barriers) (40) - Immune suppression (Tregs, MDSCs) (19, 20)	- Dense ECM prevents mAb penetration - PD-L1 overexpression suppresses the immune response	- mAbs fail to reach the tumor site - Resistance to anti-PD-1 therapy (nivolumab)	- CAR-T therapy (BBB-penetrating CAR-Ts) - Matrix-degrading enzymes for ECM remodeling

7 Future Directions and Strategies to Overcome Resistance

The growing challenge of resistance to mAb therapies has prompted the exploration of innovative strategies to enhance their efficacy and durability. These approaches leverage advancements in combination treatments, next-generation mAbs, and precision medicine to address specific resistance mechanisms. They include cutting-edge technologies like bispecific antibodies, antibody-drug conjugates (ADCs), and the integration of artificial intelligence (AI) into drug discovery and personalized treatment planning (31). Below is an in-depth exploration of these strategies:

7.1 Combination Therapies

Combination therapies use mAbs alongside other treatments, such as chemotherapy, immune checkpoint inhibitors, or small-molecule inhibitors, to achieve synergistic effects. These combinations reduce the likelihood of resistance development by simultaneously targeting multiple pathways or mechanisms. For instance, in HER2-positive breast cancer, trastuzumab is often combined with pertuzumab, another HER2-targeting antibody, to block different domains of the receptor (25). Clinical trials demonstrated that this combination improved progression-free survival by 6.1 months compared to trastuzumab alone (9). Similarly,

mAbs like cetuximab (anti-EGFR) are paired with chemotherapy for colorectal cancer to enhance tumor regression rates. These combinations are particularly effective in tumors exhibiting pathway redundancy or compensatory signaling.

7.2 Bispecific Antibodies and Their Therapeutic Potential

Bispecific antibodies (BsAbs) are a groundbreaking advancement in cancer immunotherapy, engineered to simultaneously bind two distinct targets, thereby enhancing treatment specificity and efficacy. One prominent example is blinatumomab, a bispecific T-cell engager (BiTE) that bridges CD19 on B-cell malignancies and CD3 on T-cells, promoting immune-mediated tumor killing. Clinical trials have demonstrated its remarkable efficacy, achieving a complete remission rate of 43% in patients with relapsed or refractory acute lymphoblastic leukemia (ALL) (104). Beyond BiTEs, bulbs are being developed to target complex resistance mechanisms. For instance, bulbs designed to bind HER2 and PD-L1 simultaneously inhibit tumor growth signals and counteract immune evasion, offering a dual approach to overcoming therapeutic resistance. By addressing multiple resistance pathways and engaging the immune system more effectively, bispecific antibodies hold immense potential as versatile tools for treating heterogeneous and resistant tumors.

7.3 Immune Checkpoint Inhibitors

Immune checkpoint inhibitors (ICIs), such as anti-PD-1 (nivolumab) and anti-PD-L1 (atezolizumab) antibodies, restore the immune system's ability to attack cancer cells. These inhibitors block inhibitory signals like PD-1/PD-L1, which cancer cells exploit to evade immune detection. In non-small cell lung cancer (NSCLC), nivolumab significantly improved overall survival, with a 41% reduction in mortality compared to chemotherapy (89). Combining ICIs with mAbs targeting tumor-specific antigens amplifies immune responses and reduces resistance by targeting intrinsic and extrinsic mechanisms.

7.4 Advancements in ADCs

Antibody-drug conjugates (ADCs) represent a novel therapeutic approach that combines the specificity of mAbs with the cytotoxic potency of chemotherapeutic agents. ADCs like trastuzumab deruxtecan and sacituzumab govitecan have shown promise in overcoming resistance in cancers such as HER2-positive breast cancer and triple-negative breast cancer (108). By selectively delivering cytotoxic payloads to cancer cells, ADCs minimize off-target effects, addressing systemic toxicity observed in traditional chemotherapy. For example, trastuzumab deruxtecan demonstrated a 61% objective response rate in HER2-positive breast cancer patients previously resistant to trastuzumab (88). This underscores the ability of ADCs to overcome resistance caused by antigen downregulation or genetic mutations.

7.5 Targeting Emerging Resistance Mechanisms

Research into the molecular underpinnings of resistance has revealed novel pathways that tumors exploit to evade mAb therapies. For instance, mutations in the HER2 gene can alter the receptor's conformation, diminishing the efficacy of HER2-targeted therapies like trastuzumab. Similarly, tumors frequently activate bypass pathways involving RTKs such as MET or AXL to maintain oncogenic signaling in the presence of

mAb inhibition (76). Next-generation mAbs are being designed to target multiple pathways simultaneously, reducing the likelihood of resistance emergence (98).

7.6 Nanoparticle Delivery Systems

Nanoparticles offer an innovative solution to drug delivery challenges in the tumor microenvironment (TME), which often limits the efficacy of mAb therapies. These nanocarriers encapsulate mAbs and deliver them directly to tumors, bypassing barriers such as dense stroma, high interstitial fluid pressure, and irregular vasculature. For instance, liposomal nanoparticles loaded with trastuzumab have shown improved penetration into HER2-positive breast cancer tissues, enhancing therapeutic outcomes (6). Nanoparticles can also be designed to respond to specific triggers in the TME, such as acidic pH or enzymes, releasing their payload only in the tumor site. Additionally, nanoparticle delivery systems can co-encapsulate mAbs with other drugs, such as chemotherapeutic agents or immune modulators, enabling combination therapies with precise tumor targeting and reduced systemic toxicity. For example, nanoparticles carrying trastuzumab and paclitaxel have synergistic effects in preclinical studies, offering a promising avenue for overcoming resistance and minimizing side effects. The use of nanotechnology in mAb therapies is rapidly evolving, with ongoing clinical trials exploring its potential in diverse cancer types.

7.7 Next-Generation Monoclonal Antibodies: Leveraging Protein Dynamics to Overcome Resistance

Proteins are dynamic ensembles, with binding site shapes influenced by hinge-bending motions and conformational shifts. Designing mAbs that target multiple binding site conformations can minimize resistance caused by ligand-induced population shifts and protein mutations. Building on this, next-generation mAbs, including antibody-drug conjugates (ADCs) and bispecific antibodies, are designed to overcome traditional resistance mechanisms and enhance therapeutic precision (109).

ADCs, such as trastuzumab emtansine (T-DM1), combine the targeting precision of mAbs with cytotoxic agents to kill resistant cancer cells. They achieve a 9.6-month progression-free survival advantage in HER2-positive breast cancer—bispecific antibodies further address tumor heterogeneity by simultaneously targeting two antigens, enhancing immune activation (48).

Glycoengineering advancements improve Fc receptor binding and antibody-dependent cellular cytotoxicity (ADCC), amplifying anti-tumor effects. Additionally, site-specific conjugation in ADCs and AI-driven mAb
design enhance therapeutic efficacy and reduce immunogenicity. By leveraging these innovations and protein flexibility, next-generation mAbs address resistance with precision, solidifying their role in modern oncology.

7.8 Real-Time Adaptive mAb Therapies

Real-time adaptive therapies involve a dynamic approach to counteract resistance in mAb treatments by tailoring strategies to tumor evolution. Unlike traditional static regimens, this method employs multiple mAbs targeting distinct binding pockets or epitopes of the same antigen, administered sequentially or as a cocktail. The goal is to reduce selective pressure on any binding site, delaying resistance development. For instance, in HER2-positive breast cancer, combining or alternating trastuzumab with pertuzumab, which targets a different HER2 domain, has shown improved outcomes in clinical settings (10). This approach exploits the biological diversity of tumor cells to limit their adaptive capabilities.

Technological advancements, such as liquid biopsies and next-generation sequencing (NGS), enable real-time monitoring of tumor mutations, guiding timely modifications to therapeutic regimens. For example, in colorectal cancer, detecting KRAS mutations through liquid biopsies in patients receiving cetuximab can prompt a switch to therapies targeting bypass pathways like MET or AXL. This flexibility helps address resistance mechanisms as they emerge.

Adaptive strategies include engineering mAbs with variable binding affinities or glycosylation profiles to enhance immune activation or improve binding efficiency. For example, mAbs designed with higher affinity for Fc receptors can increase antibody-dependent cellular cytotoxicity (ADCC), maintaining immune pressure on tumor cells. However, challenges remain, including regulatory hurdles, higher production costs, and the need for advanced computational models to predict tumor evolution.

Ongoing research and innovation, such as rapid mAb synthesis and AI-driven modeling, are expected to make real-time adaptive therapies more feasible and cost-effective. This approach represents a paradigm shift in oncology, focusing on preemptive and flexible treatment strategies to outpace tumor evolution and resistance.

7.9 Personalized Medicine in mAb Therapies

Personalized medicine revolutionizes mAb therapies by tailoring treatments to the unique molecular profile of each patient's tumor. This precision-based approach relies on advanced genomic, transcriptomic, and proteomic analyses to identify actionable mutations, signaling pathway alterations, or resistance mechanisms. By understanding the specific biological drivers of a tumor, clinicians can optimize therapy selection, improving efficacy and reducing unnecessary side effects. For example, in HER2-positive breast cancer, genomic profiling has guided the use of trastuzumab deruxtecan in patients resistant to trastuzumab. This therapy achieved a 61% objective response rate in resistant cases by targeting HER2-overexpressing cells while delivering a cytotoxic payload. Similarly, in non-small cell lung cancer (NSCLC), proteomic profiling has identified high PD-L1 expression, prompting anti-PD-L1 therapies like atezolizumab, which extended overall survival by 4.2 months compared to standard chemotherapy (34).

Artificial intelligence (AI) and machine learning (ML) are critical in integrating multi-omics data, predicting

resistance mechanisms, and personalizing treatment plans. AI-driven algorithms can analyze large datasets to identify biomarkers such as MET amplification or PTEN loss, which guide the use of combination therapies or next-generation mAbs. For example, AI models have accurately predicted responses to bispecific antibodies targeting HER2 and HER3, enabling precise therapy allocation(28).

Personalized medicine also influences clinical trial designs, transitioning from traditional approaches to adaptive trials. These trials use real-time molecular data to assign patients to the most effective treatment arms, improving trial outcomes and patient care. Initiatives like the National Cancer Institute's MATCH trial exemplify this strategy by matching treatments to specific genetic alterations across cancer types.

While personalized medicine faces challenges, including high costs, limited accessibility to advanced diagnostics, and the complexity of integrating diverse datasets, global initiatives are addressing these barriers. Programs like the Cancer Moonshot and the increased adoption of open-access data platforms aim to make precision oncology more accessible and practical.

8 Conclusion

The development and application of monoclonal antibody therapies have revolutionized cancer treatment, offering precision and efficacy in targeting specific molecular pathways. However, the emergence of resistance remains a formidable challenge, diminishing the long-term effectiveness of these therapies. Resistance mechanisms, driven by molecular, cellular, and tumor microenvironmental factors, underscore the complexity and adaptability of cancer biology. To sustain the therapeutic benefits of mAbs, it is crucial to address these resistance mechanisms proactively and develop innovative strategies for their prevention and management.

Resistance arises at the molecular level through antigen mutations, target expression loss, and the activation of compensatory pathways. Targeting these mechanisms requires advancements in antibody engineering, such as bispecific antibodies that simultaneously bind multiple targets or next-generation antibody-drug conjugates (ADCs) that deliver cytotoxic payloads directly to tumor cells, bypassing resistance associated with antigen loss. Incorporating techniques like glycoengineering to enhance Fc receptor binding and antibody-dependent cellular cytotoxicity (ADCC) can further optimize mAb efficacy.

The tumor microenvironment (TME) presents another critical hurdle, with immunosuppressive cells, extracellular matrix components, and hypoxia forming physical and biochemical barriers that limit mAb penetration and activity. Combination therapies are a promising avenue to counteract these barriers. Pairing mAbs with agents that remodel the TME, such as anti-angiogenic drugs or stromal-depleting agents, can enhance antibody delivery and efficacy. Nanoparticle-based delivery systems also offer potential, allowing mAbs to navigate the dense and hypoxic TME more effectively while minimizing systemic toxicity. Advances in genomic and proteomic profiling are paving the way for precision medicine, enabling personalized treatment plans tailored to individual resistance mechanisms. Liquid biopsies and next-generation sequencing can help identify emerging resistance mutations or alterations, allowing for real-time therapeutic adjustments.

This adaptive approach ensures that patients receive the most effective therapies while mitigating the risk of resistance development.

Furthermore, addressing immune evasion mechanisms, such as the upregulation of immune checkpoints like PD-L1, is essential. Combining immune checkpoint inhibitors with mAbs targeting specific tumor antigens can restore anti-tumor immune responses while enhancing the overall efficacy of therapy. Strategies to modulate the immune system, such as cytokine therapies or vaccines, also promise to overcome immune suppression and reinvigorate immune surveillance.

Looking ahead, leveraging artificial intelligence (AI) and machine learning can accelerate drug discovery, optimize treatment combinations, and predict resistance patterns. By integrating data from genomics, proteomics, and clinical outcomes, AI-driven models can guide the development of novel mAbs and therapeutic regimens with greater precision.

Abbreviations

ADC(s): Antibody-drug conjugate(s)

ALL: Acute lymphoblastic leukemia

BiTE(s): Bispecific T-cell engager(s)

CML: Chronic myelogenous leukemia

CRC: Colorectal cancer

EGFR: Epidermal growth factor receptor

GIST: Gastrointestinal stromal tumors

HGF: Hepatocyte growth factor

ICI(s): Immune checkpoint inhibitor(s)

mAb(s): Monoclonal antibody

(antibodies) **MET:** Hepatocyte growth

factor receptor **NSCLC:** Non-small cell

lung carcinoma **NRTK(s):** Non-receptor

tyrosine kinase(s) **PD-1:** Programmed

death-1

PD-L1: Programmed death-ligand 1

RTK(s): Receptor tyrosine kinase(s)

TGFα: Transforming growth factor alpha

TKI(s): Tyrosine kinase inhibitor(s)

HER2: Human epidermal growth factor receptor 2

References

(1) Adler, M. J., & Dimitrov, D. S. (2012). Therapeutic antibodies against cancer. Hematology/Oncology Clinics of North America, 26(3), 447–481. https://pmc.ncbi.nlm.nih.gov/articles/PMC3334873/ (2) American Cancer Society. (2023, December 20). Monoclonal antibodies and their side effects. American CancerSociety.https://www.cancer.org/cancer/managing-cancer/treatment-types/immunotherapy/monoclonal-antibodies.html

(3) Ayob, A. Z., & Ramasamy, T. S. (2018). Cancer stem cells as key drivers of tumour progression. Journal of Biomedical Science, 25(1), 20. https://doi.org/10.1186/s12929-018-0426-4

(4) Baez Rodriguez, S. M., Kamel, A., Ciubotaru, G. V., Onose, G., Sevastre, A.-S., & Tataranu, L. G. (2023) etc. An overview of EGFR mechanisms and their implications in targeted therapies for glioblastoma. International Journal of Molecular Sciences, 24(13), 11110. https://doi.org/10.3390/ijms241311110

(5) Bahar, M. E., Kim, H. J., & Kim, D. R. (2023). Targeting the RAS/RAF/MAPK pathway for cancer therapy: From mechanism to clinical studies. Signal Transduction and Targeted Therapy, 8, Article 455. https://doi.org/10.1038/s41392-023-01705-z

(6) Bahar, M. E., Kim, H. J., & Kim, D. R. (2023). Targeting the RAS/RAF/MAPK pathway for cancer therapy: From mechanism to clinical studies. Signal Transduction and Targeted Therapy, 8, Article 455. https://doi.org/10.1038/s41392-023-01705-z

(7) Bakema, J. E., & van Egmond, M. (2014). Fc receptor-dependent mechanisms of monoclonal antibody therapy of cancer. In M. W. Wels (Ed.), Molecular Oncology: Causes of Cancer and Targets for Treatment (pp. 355–374). Springer. https://doi.org/10.1007/978-3-319-07911-0_17

(8) Bardelli, A., & Siena, S. (2010). Molecular mechanisms of resistance to cetuximab and panitumumab in colorectal cancer. Journal of Clinical Oncology, 28(7), 1254–1261. https://doi.org/10.1200/JCO.2009.24.6116

(9) Baselga, J., Carbonell, X., Castañeda-Soto, N. J., Clemens, M., Green, M., Harvey, V., et al. (2005). Phase II study of efficacy, safety, and pharmacokinetics of trastuzumab monotherapy administered on a 3-weekly schedule. Journal of Clinical Oncology, 23(10), 2162–2171. https://doi.org/10.1200/JCO.2005.01.014

(10) Baselga, J., Carbonell, X., Castañeda-Soto, N. J., Clemens, M., Green, M., Harvey, V., et al. (2005). Phase

II study of efficacy, safety, and pharmacokinetics of trastuzumab monotherapy administered on a 3-weekly schedule. Journal of Clinical Oncology, 23(10), 2162–2171. https://doi.org/10.1200/JCO.2005.01.014

(11) Beck, A., Goetsch, L., Dumontet, C., & Corvaïa, N. (2017). Strategies and challenges for the next generation of antibody-drug conjugates. Nature Reviews Drug Discovery, 16(5), 315–337. https://doi.org/10.1038/nrd.2016.268

(12) Bernard-Marty, C., Lebrun, F., Awada, A., & Piccart, M. J. (2006). Monoclonal antibody-based targeted therapy in breast cancer: Current status and future directions. Drugs, 66(12), 1577–1591. https://doi.org/10.2165/00003495-200666120-00004

(13) Boussiotis, V. A. (2016). Molecular and biochemical aspects of the PD-1 checkpoint pathway. The New England Journal of Medicine, 375(18), 1767–1778. https://doi.org/10.1056/NEJMra1514296

(14) Boussiotis, V. A. (2016). Molecular and biochemical aspects of the PD-1 checkpoint pathway. The New England Journal of Medicine, 375(18), 1767–1778. https://doi.org/10.1056/NEJMra1514296

(15) Callaghan, R., Luk, F., & Bebawy, M. (2014). Inhibition of the multidrug resistance P-glycoprotein: Time for a change of strategy? Drug Metabolism and Disposition, 42(4), 623–631. https://doi.org/10.1124/dmd.113.056176

(16) Cancer Research UK. (2021, January 8). Monoclonal antibodies (MABs). Cancer Research UK. https://www.cancerresearchuk.org/about-cancer/treatment/targeted-cancer-drugs/types/monoclonal-antibodies

(17) Chames, P., Van Regenmortel, M., Weiss, E., & Baty, D. (2009). Therapeutic antibodies: Successes, limitations, and hopes for the future. British Journal of Pharmacology, 157(2), 220–233. https://doi.org/10.1111/j.1476-5381.2009.00190.x

(18) Chames, P., Van Regenmortel, M., Weiss, E., & Baty, D. (2009). Therapeutic antibodies: Successes, limitations, and hopes for the future. British Journal of Pharmacology, 157(2), 220–233. https://doi.org/10.1111/j.1476-5381.2009.00190.x

(19) Chaudhary, B., & Elkord, E. (2016). Regulatory T cells in the tumor microenvironment and cancer progression: Role and therapeutic targeting. Frontiers in Immunology, 7, Article 183. https://doi.org/10.3389/fimmu.2016.00183

(20) Chaudhary, B., & Elkord, E. (2016). Regulatory T cells in the tumor microenvironment and cancer progression: Role and therapeutic targeting. Frontiers in Immunology, 7, Article 183. https://doi.org/10.3389/fimmu.2016.00183

(21) Chaudhary, B., & Elkord, E. (2016). Regulatory T cells in the tumor microenvironment and cancer progression: Role and therapeutic targeting. Frontiers in Immunology, 7, Article 183. https://doi.org/10.3389/fimmu.2016.00183

(22) Chen, J., Guo, Y., Zeng, W., Huang, Y., & Zhao, M. (2019). Metabolic reprogramming in solid tumors: targeting OXPHOS and glycolysis with mAbs. Biomolecules, 10(5), 701. https://doi.org/10.3390/biom10050701

(23) Cheng, L., Chen, L., Shi, Y., Gu, W., Ding, W., Zheng, X., Liu, Y., Jiang, J., & Zheng, Z. (2024). Efficacy and safety of bispecific antibodies vs. immune checkpoint blockade combination therapy in cancer: A real-world comparison. Molecular Cancer, 23, Article 77. https://doi.org/10.1186/s12943-024-01956-6

(24) Cheng, L., Chen, L., Shi, Y., Gu, W., Ding, W., Zheng, X., Liu, Y., Jiang, J., & Zheng, Z. (2024). Efficacy and safety of bispecific antibodies vs. immune checkpoint blockade combination therapy in cancer: A real-world comparison. Molecular Cancer, 23, Article 77. https://doi.org/10.1186/s12943-024-01956-6

(25) Cheng, L., Chen, L., Shi, Y., Gu, W., Ding, W., Zheng, X., Liu, Y., Jiang, J., & Zheng, Z. (2024). Efficacy and safety of bispecific antibodies vs. immune checkpoint blockade combination therapy in cancer: A real-world comparison. Molecular Cancer, 23, Article 77. https://doi.org/10.1186/s12943-024-01956-6

(26) Day, C.-P., Merlino, G., & Van Dyke, T. (2015). Preclinical mouse cancer models: A maze of opportunities and challenges. Cell, 163(1), 39–53. https://doi.org/10.1016/j.cell.2015.08.068

(27) Day, C.-P., Merlino, G., & Van Dyke, T. (2015). Preclinical mouse cancer models: A maze of opportunities and challenges. Cell, 163(1), 39–53. https://doi.org/10.1016/j.cell.2015.08.068

(28) Dhudum, R., Ganeshpurkar, A., & Pawar, A. (2024). Revolutionizing drug discovery: A comprehensive review of AI applications. Drugs Drug Candidates, 3(1), 148–171. https://doi.org/10.3390/ddc3010009

(29) Fang, L., He, R., Cai, W., & Huang, D. (2024). Mapping the mutational landscape of immunotherapy-responsive cancers. Nature Genetics, 56(3), 312–325. https://doi.org/10.1038/s41588-024-01948-8

(30) Ghemrawi, R., Abuamer, L., Kremesh, S., Hussien, G., Ahmed, R., Mousa, W., Khoder, G., & Khair, M. (2024). Revolutionizing cancer treatment: Recent advances in immunotherapy. Biomedicines, 12(9), 2158. https://doi.org/10.3390/biomedicines12092158

(31) Ghemrawi, R., Abuamer, L., Kremesh, S., Hussien, G., Ahmed, R., Mousa, W., Khoder, G., & Khair, M. (2024). Revolutionizing cancer treatment: Recent advances in immunotherapy. Biomedicines, 12(9), 2158. https://doi.org/10.3390/biomedicines12092158

(32) Ghemrawi, R., Abuamer, L., Kremesh, S., Hussien, G., Ahmed, R., Mousa, W., Khoder, G., & Khair, M. (2024). Revolutionizing cancer treatment: Recent advances in immunotherapy. Biomedicines, 12(9), 2158. https://doi.org/10.3390/biomedicines12092158

(33) Gonzalez de Castro, D., Clarke, P. A., Al-Lazikani, B., & Workman, P. (2013). Personalized cancer

medicine: Molecular diagnostics, predictive biomarkers, and drug resistance. Clinical Pharmacology & Therapeutics, 93(3), 252–259. https://doi.org/10.1038/clpt.2012.237

(34) Gonzalez de Castro, D., Clarke, P. A., Al-Lazikani, B., & Workman, P. (2013). Personalized cancer medicine: Molecular diagnostics, predictive biomarkers, and drug resistance. Clinical Pharmacology & Therapeutics, 93(3), 252–259. https://doi.org/10.1038/clpt.2012.237

(35) Gravbrot, N., Gilbert-Gard, K., Mehta, P., Ghotmi, Y., Banerjee, M., Mazis, C., & Sundararajan, S. (2019). Therapeutic monoclonal antibodies targeting immune checkpoints for the treatment of solid tumors. Journal of Immunotherapy and Precision Oncology, 8(4), 215–225. https://doi.org/10.3389/fonc.2019.00763

(36) Haibe, Y., Kreidieh, M., El Hajj, H., Khalifeh, I., Mukherji, D., Temraz, S., & Shamseddine, A. (2020). Resistance mechanisms to anti-angiogenic therapies in cancer. Frontiers in Oncology, 10, 221. https://doi.org/10.3389/fonc.2020.00221

(37) Hackstadt, T. (1988). Steric hindrance of antibody binding to surface proteins of Coxiella burnetii by phase I lipopolysaccharide. Infection and Immunity, 56(4), 802–807. https://doi.org/10.1128/iai.56.4.802-807.1988

(38) Hansel, T. T., Kropshofer, H., Singer, T., Mitchell, J. A., & George, A. J. T. (2010). The safety and side effects of monoclonal antibodies. Nature Reviews Drug Discovery, 9(4), 325–338. https://doi.org/10.1038/nrd3003

(39) Heady, D. (2023, October 18). 25 years of Herceptin: A groundbreaking advancement in breast cancer treatment. UCLA Health News. https://www.uclahealth.org/news/release/25-years-herceptin-groundbreaking-advancement-breast-cancer

(40) Henke, E., Nandigama, R., & Ergün, S. (2020). Extracellular matrix in the tumor microenvironment and its impact on cancer therapy. Frontiers in Molecular Biosciences, 6(160). https://doi.org/10.3389/fmolb.2019.00160

(41) Horvath, L., Thienpont, B., Zhao, L., Wolf, D., & Pircher, A. (2020). Overcoming immunotherapy resistance in non-small cell lung cancer (NSCLC) - novel approaches and future outlook. Molecular Cancer, 19, Article 141. https://doi.org/10.1186/s12943-020-01260-z

(42) Horvath, L., Thienpont, B., Zhao, L., Wolf, D., & Pircher, A. (2020). Overcoming immunotherapy resistance in non-small cell lung cancer (NSCLC) - novel approaches and future outlook. Molecular Cancer, 19, Article 141. https://doi.org/10.1186/s12943-020-01260-z

(43) Hosseinzadeh, A., Merikhian, P., Naseri, N., Eisavand, M. R., & Farahmand, L. (2022). MUC1 is a potential target to overcome trastuzumab resistance in breast cancer therapy. Cancer Cell International, 22, 110. https://doi.org/10.1186/s12935-022-02523-z

(44) International Agency for Research on Cancer (IARC). (2020, December 15). Latest global cancer data: Cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020. IARC Press Release No. 292. Retrieved from
https://www.iarc.who.int/news-events/latest-global-cancer-data-cancer-burden-rises-to-19-3-million-new-cases and-10-0-million-cancer-deaths-in-2020/

(45) Iqbal, N., & Iqbal, N. (2014). Human epidermal growth factor receptor 2 (HER2) in cancers: Overexpression and therapeutic implications. Molecular Biology International, 2014, 852748. https://doi.org/10.1155/2014/852748

(46) Jones, P. A., & Baylin, S. B. (2007). The epigenomics of cancer. Cell, 128(4), 683–692. https://doi.org/10.1016/j.cell.2007.01.029

(47) Kawaguchi, Y., Kono, K., Mimura, K., Mitsui, F., Sugai, H., Akaike, H., & Fujii, H. (2007). Targeting EGFR and HER-2 with cetuximab- and trastuzumab-mediated immunotherapy in oesophageal squamous cell carcinoma. British Journal of Cancer, 97(4), 494–501. https://doi.org/10.1038/sj.bjc.6603901

(48) Kontermann, R. E. (2012). Dual targeting strategies with bispecific antibodies. MAbs, 4(2), 182–197. https://doi.org/10.4161/mabs.4.2.19000

(49) Lee, Y. P., Lee, M. S., Kim, H. S., Kim, J. Y., Ahn, J. S., Im, Y. H., & Park, Y. H. (2022). Real-world evidence of trastuzumab, pertuzumab, and docetaxel combination as a first-line treatment for Korean patients with HER2-positive metastatic breast cancer. Cancer Research and Treatment, 54(4), 1130–1137. https://doi.org/10.4143/crt.2021.1103

(50) Li, T., Niu, M., Zhou, J., Wu, K., & Yi, M. (2024). The enhanced antitumor activity of bispecific antibody targeting PD-1/PD-L1 signaling. Cell Communication and Signaling, 22, Article 179. https://doi.org/10.1186/s12964-024-01562-5

(51) Li, X., Xu, W., Jiang, H., & Wang, Y. (2024). Targeting amino acid transporters in cancer immunotherapy. Acta Pharmaceutica Sinica B. https://doi.org/10.1016/j.apsb.2024.07.024

(52) Liu, B., Zhou, H., Tan, L., Siu, K. T. H., & Guan, X.-Y. (2024). Exploring treatment options in cancer: Tumor treatment strategies. Signal Transduction and Targeted Therapy, 9(1), 175. https://doi.org/10.1038/s41392-024-01856-7

(53) Liu, R. M., Hwang, Y. C., Liu, I. J., Lee, C. C., Tsai, H. Z., Li, H. J., & Wu, H. C. (2020). Development of therapeutic antibodies for the treatment of diseases. Journal of Biomedical Science, 27(1), Article 1. https://doi.org/10.1186/s12929-019-0592-z

(54) Liu, R. M., Hwang, Y. C., Liu, I. J., Lee, C. C., Tsai, H. Z., Li, H. J., & Wu, H. C. (2020). Development of therapeutic antibodies for the treatment of diseases. Journal of Biomedical Science, 27(1), Article 1.

https://doi.org/10.1186/s12929-019-0592-z

(55) Liu, R., Hwang, Y., Liu, I., Lee, C., Tsai, H., Li, H., & Wu, H. (2020). Development of therapeutic antibodies for the treatment of diseases. Journal of Biomedical Science, 27(1).
https://doi.org/10.1186/s12929-019-0592-z

(56) Liu, Z. L., Chen, H. H., Zheng, L. L., Sun, L. P., & Shi, L. (2023). Angiogenic signaling pathways and anti-angiogenic therapy for cancer. Signal Transduction and Targeted Therapy, 8, Article 198.
https://doi.org/10.1038/s41392-023-01460-1

(57) Loganzo, F., Sung, M., & Gerber, H.-P. (2016). Mechanisms of resistance to antibody–drug conjugates. Molecular Cancer Therapeutics, 15(12), 2825–2834.
https://doi.org/10.1158/1535-7163.MCT-16-0408

(58) Mahgoub, E. O., Cho, W. C., Sharifi, M., Falahati, M., Zeinabad, H. A., Mare, H. E., & Hasan, A. (2023). Role of functional genomics in identifying cancer drug resistance and overcoming cancer relapse. Heliyon, 10(1), e22095. https://doi.org/10.1016/j.heliyon.2023.e22095

(59) Mak, K. K., & Pichika, M. R. (2019). Artificial intelligence in drug development: Present status and prospects. Drug Discovery Today, 24(3), 773–780. https://doi.org/10.1016/j.drudis.2018.11.014

(60) Maruyama, I. N. (2014). Mechanisms of activation of receptor tyrosine kinases: Monomers or dimers. Cells, 3(2), 304–330. https://doi.org/10.3390/cells3020304

(61) Moon, D., Tae, N., Park, Y., Lee, S.-W., & Kim, D. H. (2022). Development of bispecific antibody for cancer immunotherapy: Focus on T cell engaging antibody. Immune Network, 22(1), e4.
https://doi.org/10.4110/in.2022.22.e4

(62) Nahta, R., Yu, D., Hung, M. C., Hortobagyi, G. N., & Esteva, F. J. (2006). Mechanisms of disease: Understanding resistance to HER2-targeted therapy in human breast cancer. Nature Clinical Practice Oncology, 3(5), 269–280. https://doi.org/10.1038/ncponc0509

(63) Nahta, R., Yu, D., Hung, M. C., Hortobagyi, G. N., & Esteva, F. J. (2006). Mechanisms of disease: Understanding resistance to HER2-targeted therapy in human breast cancer. Nature Clinical Practice Oncology, 3(5), 269–280. https://doi.org/10.1038/ncponc0509

(64) Nahta, R., Yu, D., Hung, M. C., Hortobagyi, G. N., & Esteva, F. J. (2006). Mechanisms of disease: Understanding resistance to HER2-targeted therapy in human breast cancer. Nature Clinical Practice Oncology, 3(5), 269–280. https://doi.org/10.1038/ncponc0509

(65) Nahta, R., Yu, D., Hung, M. C., Hortobagyi, G. N., & Esteva, F. J. (2006). Mechanisms of disease: Understanding resistance to HER2-targeted therapy in human breast cancer. Nature Clinical Practice Oncology, 3(5), 269–280. https://doi.org/10.1038/ncponc0509

(66) Niederst, M. J., & Engelman, J. A. (2013). Bypass mechanisms of resistance to receptor tyrosine kinase inhibition in lung cancer. Science Signaling, 6(294), ra25. https://doi.org/10.1126/scisignal.2004652

(67) Niederst, M. J., & Engelman, J. A. (2013). Bypass mechanisms of resistance to receptor tyrosine kinase inhibition in lung cancer. Science Signaling, 6(294), ra25. https://doi.org/10.1126/scisignal.2004652

(68) Niederst, M. J., & Engelman, J. A. (2013). Bypass mechanisms of resistance to receptor tyrosine kinase inhibition in lung cancer. Science Signaling, 6(294), re6. https://pmc.ncbi.nlm.nih.gov/articles/PMC3876281/

(69) Olson, B. M., & McNeel, D. G. (2013). Antigen loss and tumor-mediated immunosuppression facilitate tumor recurrence. Cancer Immunology, Immunotherapy, 62(7), 1123–1135. https://doi.org/10.1007/s00262-012-1387-9

(70) Paul, M. K., & Mukhopadhyay, A. K. (2004). Tyrosine kinase – Role and significance in cancer. International Journal of Medical Sciences, 1(2), 101–115. https://doi.org/10.7150/ijms.1.101

(71) Pradeep, S. P., & Bahal, R. (2024, November 22). Advancements in cell-penetrating monoclonal antibody treatment. Oncotarget, 15, 815–816. https://doi.org/10.18632/oncotarget.28674

(72) Quinteros, D. A., Bermúdez, J. M., Ravetti, S., Cid, A., Allemandi, D. A., & Palma, S. D. (2017). Therapeutic use of monoclonal antibodies: General aspects and challenges for drug delivery. In E. Andronescu & A. M. Grumezescu (Eds.), Nanostructures for Drug Delivery (pp. 807–833). Elsevier. https://doi.org/10.1016/B978-0-323-46143-6.00025-7

(73) Rapid Novor. (2023). Antibody glycan engineering: increasing potency by afucosylation. RapidNovor Knowledge Hub. https://www.rapidnovor.com/antibody-glycan-engineering/

(74) Rodríguez-Nava, C., Ortuño-Pineda, C., Illades-Aguiar, B., Flores-Alfaro, E., Leyva-Vázquez, M. A., Parra-Rojas, I., et al. (2023). Mechanisms of action and limitations of monoclonal antibodies and single chain fragment variable (scFv) in cancer treatment. Biomedicines, 11(6), 1610. https://doi.org/10.3390/biomedicines11061610

(75) Rodríguez-Nava, C., Ortuño-Pineda, C., Illades-Aguiar, B., Flores-Alfaro, E., Leyva-Vázquez, M. A., Parra-Rojas, I., Del Moral-Hernández, O., Vences-Velázquez, A., Cortés-Sarabia, K., & Alarcón-Romero, L. D. C. (2023). Mechanisms of Action and Limitations of Monoclonal Antibodies and Single Chain Fragment Variable (scFv) in the Treatment of Cancer. Biomedicines, 11(6), 1610. https://doi.org/10.3390/biomedicines11061610

(76) Roskoski, R., Jr. (2014). The ErbB/HER family of protein-tyrosine kinases and cancer. Pharmacological Research, 79, 34–74. https://doi.org/10.1016/j.phrs.2013.11.002

(77) Sadida, H. Q., Abdulla, A., Al Marzooqi, S., Hashem, S., Macha, M. A., Al-Shabeeb Akil, A. S., & Bhat, A. A. (2023). Epigenetic modifications: Key players in cancer heterogeneity and drug resistance. Translational Oncology, 39, 101821. https://doi.org/10.1016/j.tranon.2023.101821

(78) Salles, G., Barrett, M., Foà, R., Maurer, J., O'Brien, S., Valente, N., Wenger, M., & Maloney, D. G. (2017). Rituximab in B-cell hematologic malignancies: A review of 20 years of clinical experience. Blood Advances, 1(20), 2388–2406. https://doi.org/10.1182/bloodadvances.2017007903

(79) Shiravand, Y., Khodadadi, F., Kashani, S. M. A., Hosseini-Fard, S. R., Hosseini, S., Sadeghirad, & Kulasinghe, A. (2022). Immune checkpoint inhibitors in cancer therapy. Frontiers in Immunology, 13, Article 912391. https://doi.org/10.3389/fimmu.2022.912391

(80) Shiravand, Y., Khodadadi, F., Kashani, S. M. A., Hosseini-Fard, S. R., Hosseini, S., Sadeghirad, & Kulasinghe, A. (2022). Immune checkpoint inhibitors in cancer therapy. Frontiers in Immunology, 13, Article 912391. https://doi.org/10.3389/fimmu.2022.912391

(81) Shuptrine, C., Surana, R., & Weiner, L. M. (2012). Monoclonal antibodies for the treatment of cancer. Seminars in Cancer Biology, 22(1), 3–13. https://doi.org/10.1016/j.semcancer.2011.12.009

(82) Shuptrine, C., Surana, R., & Weiner, L. M. (2012). Monoclonal antibodies for the treatment of cancer. Seminars in Cancer Biology, 22(1), 3–13. https://doi.org/10.1016/j.semcancer.2011.12.009

(83) Shuptrine, C., Surana, R., & Weiner, L. M. (2012). Monoclonal antibodies for the treatment of cancer. Seminars in Cancer Biology, 22(1), 3–13. https://doi.org/10.1016/j.semcancer.2011.12.009

(84) Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., et al. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England Journal of Medicine, 344(11), 783–792. https://doi.org/10.1056/NEJM200103153441101

(85) Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., et al. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England Journal of Medicine, 344(11), 783–792. https://doi.org/10.1056/NEJM200103153441101

(86) Sledge, G. W., Mamounas, E. P., Hortobagyi, G. N., Burstein, H. J., Goodwin, P. J., & Wolff, A. C. (2014). Past, present, and future challenges in breast cancer treatment. Journal of Clinical Oncology, 32(19), 1979–1986. https://doi.org/10.1200/JCO.2014.55.4139

(87) Tonon, G., Rizzolio, F., Visentin, F., & Scattolin, T. (2024). Antibody-drug conjugates for cancer therapy: From metallodrugs to nature-inspired payloads. International Journal of Molecular Sciences, 25(16), 8651. https://doi.org/10.3390/ijms25168651

(88) Tonon, G., Rizzolio, F., Visentin, F., & Scattolin, T. (2024). Antibody drug conjugates for cancer therapy: From metallodrugs to nature-inspired payloads. International Journal of Molecular Sciences, 25(16), 8651. https://doi.org/10.3390/ijms25168651

(89) Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. The New England Journal of Medicine, 366(26), 2443–2454. https://doi.org/10.1056/NEJMoa1200690

(90) Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

(91) Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

(92) Tsao, L. C., Force, J., & Hartman, Z. C. (2021). Mechanisms of therapeutic anti-tumor monoclonal antibodies. Frontiers in Oncology, 11(PMCID: PMC8448950). https://doi.org/10.3389/fonc.2021.673636

(93) Tsao, L. C., Force, J., & Hartman, Z. C. (2021). Mechanisms of therapeutic anti-tumor monoclonal antibodies. Frontiers in Oncology, 11(PMCID: PMC8448950). https://doi.org/10.3389/fonc.2021.673636.

(94) Tsao, L. C., Force, J., & Hartman, Z. C. (2021). Mechanisms of therapeutic anti-tumor monoclonal antibodies. Frontiers in Oncology, 11(PMCID: PMC8448950). https://doi.org/10.3389/fonc.2021.673636

(95) Valabrega, G., Montemurro, F., & Aglietta, M. (2007). Trastuzumab: Mechanism of action, resistance, and future perspectives in HER2-overexpressing breast cancer. Annals of Oncology, 18(6), 977–984. https://doi.org/10.1093/annonc/mdl475

(96) Van der Horst, H. J., & Mutis, T. (2024). Enhancing Fc-mediated effector functions of monoclonal antibodies: The example of HexaBodies. Immunological Reviews. Advance online publication. https://doi.org/10.1111/imr.13394

(97) Van der Horst, H. J., Nijhof, I. S., Mutis, T., & Chamuleau, M. E. D. (2020). Fc-engineered antibodies with enhanced Fc-effector function for the treatment of B-cell malignancies. Cancers (Basel), 12(10), 3041. https://doi.org/10.3390/cancers12103041

(98) Waks, A. G., & Winer, E. P. (2019). Breast cancer treatment: A review. JAMA, 321(3), 288–300. https://doi.org/10.1001/jama.2018.19323

(99) Waldman, A. D., Fritz, J. M., & Lenardo, M. J. (2020). A guide to cancer immunotherapy: From T cell basic science to clinical practice. Nature Reviews Immunology, 20(11), 651–668. https://doi.org/10.1038/s41577-020-0306-5

(100) Wang, Z. H., Zheng, Z. Q., Jia, S. C., Liu, S. N., Xiao, X. F., Chen, G. Y., & Liang, W. Q., etc. (2022). Trastuzumab resistance in HER2-positive breast cancer: Mechanisms, emerging biomarkers and targeting agents. Frontiers in Oncology, 12, 9584623. https://doi.org/10.3389/fonc.2022.9584623

(101) Wang, Z. H., Zheng, Z. Q., Jia, S. C., Liu, S. N., Xiao, X. F., Chen, G. Y., & Liang, W. Q., etc. (2022). Trastuzumab resistance in HER2-positive breast cancer: Mechanisms, emerging biomarkers and targeting agents. Frontiers in Oncology, 12, 9584623. https://doi.org/10.3389/fonc.2022.9584623

(102) World Health Organization. (n.d.). Monoclonal antibodies (mABs). World Health Organization. https://www.who.int/teams/health-product-policy-and-standards/standards-and-specifications/norms-and-standa
rds/monoclonal-antibodies

(103) Wu, Y., Yi, M., Zhu, S., Wang, H., & Wu, K. (2021). Recent advances and challenges of bispecific antibodies in solid tumors. Experimental Hematology & Oncology, 10(1), Article 56. https://doi.org/10.1186/s40164-021-00250-1

(104) Wu, Y., Yi, M., Zhu, S., Wang, H., & Wu, K. (2021). Recent advances and challenges of bispecific antibodies in solid tumors. Experimental Hematology & Oncology, 10(1), Article 56. https://doi.org/10.1186/s40164-021-00250-1

(105) Xue, X., & Liang, X.-J. (2012). Overcoming drug efflux-based multidrug resistance in cancer with nanotechnology. Chinese Journal of Cancer, 31(2), 100–109. https://doi.org/10.5732/cjc.011.10326

(106) Xue, X., & Liang, X.-J. (2012). Overcoming drug efflux-based multidrug resistance in cancer with nanotechnology. Chinese Journal of Cancer, 31(2), 100–109. https://doi.org/10.5732/cjc.011.10326

(107) Xue, X., & Liang, X.-J. (2012). Overcoming drug efflux-based multidrug resistance in cancer with nanotechnology. Chinese Journal of Cancer, 31(2), 100–109. https://doi.org/10.5732/cjc.011.10326

(108) Xue, X., & Liang, X.-J. (2012). Overcoming drug efflux-based multidrug resistance in cancer with nanotechnology. Chinese Journal of Cancer, 31(2), 100–109. https://doi.org/10.5732/cjc.011.10326

(109) Xue, X., & Liang, X.-J. (2012). Overcoming drug efflux-based multidrug resistance in cancer with nanotechnology. Chinese Journal of Cancer, 31(2), 100–109. https://doi.org/10.5732/cjc.011.10326

(110) Yabroff, K. R., Lund, J., Kepka, D., & Mariotto, A. (2011). Economic burden of cancer in the US: Estimates, projections, and future research. Cancer Epidemiology, Biomarkers & Prevention, 20(10), 2006–2014. https://doi.org/10.1158/1055-9965.EPI-11-0650

(111) Yip, H. Y. K., & Papa, A. (2021). Signaling pathways in cancer: Therapeutic targets,

combinatorial treatments, and new developments. Cells, 10(3), 659.
https://doi.org/10.3390/cells10030659

(112) Zhao, Y., Yang, Y., & Wang, H. (2020). Tumor antigen heterogeneity and resistance to immune
checkpoint inhibitors. Oncotarget, 11(7), 624–637. https://doi.org/10.18632/oncotarget.15837