

# **20 COGNITIVE BIASES EVERYONE SHOULD KNOW**

+

**Bonus Chapter: Algorithmic Bias**

-

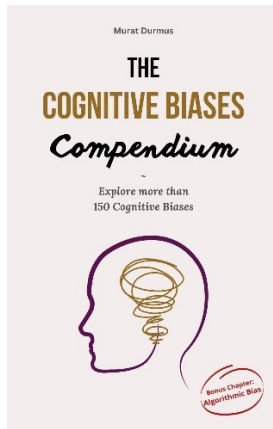
Murat Durmus

This document is an excerpt from the book:

## The Cognitive Biases Compendium

-

Explore over 160 Cognitive Biases (with examples) to make better decisions, think critically, solve problems effectively, and communicate more accurately.



Available on Amazon:

Paperback: <https://www.amazon.com/dp/B0C9SF8KZW>

eBook: <https://www.amazon.com/dp/B0CBK1WFNN>

Murat Durmus

ISBN: 9798851721496

## About the Author

Murat Durmus is CEO and founder of AISOMA (a Frankfurt am Main (Germany) based company specialized in AI-based technology development and consulting) and Author of the books "[MINDFUL AI – Reflections on Artificial Intelligence](#)".& "[A Primer to the 42 Most commonly used Machine Learning Algorithms \(With Code Samples\)](#)"

You can get in touch with the Author via:

- LinkedIn: <https://www.linkedin.com/in/ceosaisoma/>
- E-Mail: [murat.durmus@aisoma.de](mailto:murat.durmus@aisoma.de)

PREFACE .....	6
Confirmation Bias.....	8
Availability Heuristic.....	12
Anchoring or Focalism.....	15
Overconfidence Effect.....	18
Halo Effect .....	22
Recency Effect .....	27
Framing Effect .....	30
Sunk Cost Fallacy .....	34
Hindsight Bias.....	38
Loss Aversion.....	41
Gambler's Fallacy .....	45
Self-Serving Bias .....	48
Dunning-Kruger Effect.....	51
Social-Desirability Bias .....	56
Illusory Correlation.....	60
Mere-Exposure Effect .....	63
Choice-Supportive Bias .....	67
Bandwagon Effect .....	71
Negativity Bias.....	74
Fundamental Attribution Error .....	77
ALGORITHMIC BIAS .....	80
<b>More Books by the Author .....</b>	<b>97</b>

A PRIMER TO THE 42 MOST COMMONLY USED MACHINE LEARNING ALGORITHMS BOOK.....	97
THOUGHT-PROVOKING QUOTES & CONTEMPLATIONS FROM FAMOUS PSYCHOLOGISTS .....	98
MINDFUL AI <i>Reflections on Artificial Intelligence</i> ....	99
REFERENCES .....	100

# PREFACE

Cognitive biases refer to systematic patterns of deviation from normative and rational judgment. These biases are extensively studied in the fields of psychology and behavioral economics.

While many of these biases have been confirmed through reproducible research, there is an ongoing debate about how to classify and explain them. Some experts, such as Gerd Gigerenzer, criticize labeling cognitive biases as errors of judgment and argue that they can be interpreted as rational deviations from logical reasoning.

Explanations for these biases involve using information-processing rules, known as heuristics, which the brain employs to make decisions or judgments. Biases can manifest in various forms, encompassing cognitive biases driven by mental noise and motivational biases influenced by wishful thinking. Often, both types of biases coexist simultaneously.

Controversy surrounds certain biases, with debates questioning whether they are considered useless or irrational or if they contribute to positive attitudes and behavior. For instance, in social interactions, people ask

leading questions to confirm their assumptions about others. However, this confirmation bias has also been regarded as a social skill that aids in building connections.

Although much of the research on biases has been conducted with human subjects, there is evidence of biases observed in nonhumans. For instance, loss aversion has been demonstrated in monkeys, while hyperbolic discounting has been observed in rats, pigeons, and monkeys.

This book covers 20 (In the original version 168) cognitive biases, some extensively researched while others loosely understood. Nonetheless, the book aims to provide a comprehensive overview and introduction to cognitive biases. A chapter on "Algorithmic Biases" has been included, recognizing the growing significance of addressing biases in artificial intelligence systems used for decision-making.

***Let's learn more about our human biases to make less biased conclusions in the future.***

Murat Durmus

# Confirmation Bias

## Confirmation bias

Confirmation bias is a cognitive bias that affects how people seek, interpret, and recall information, often leading them to favor information that aligns with their preexisting beliefs or values. Individuals tend to exhibit confirmation bias by actively seeking out and selecting information that supports their existing views while disregarding or downplaying contradictory evidence. This bias also manifests when people interpret ambiguous information in a manner that confirms their existing attitudes.

The tendency towards confirmation bias is prevalent across various domains, including personal opinions, political ideologies, and deeply held beliefs. When individuals encounter information that aligns with their preconceived notions, it reinforces their confidence and provides a sense of validation. On the other hand, when people encounter information that contradicts their beliefs, they may feel threatened and experience cognitive dissonance. As a result, they may reject or rationalize the information.



Confirmation bias plays a significant role in shaping individuals' decision-making processes, as it can lead to a narrow focus on information that supports their desired outcomes or emotional preferences. This bias can hinder critical thinking and prevent individuals from objectively considering alternative perspectives or weighing evidence impartially.

While confirmation bias cannot be completely eliminated, awareness of its influence and deliberate efforts to manage it can mitigate its impact. Education and training in critical thinking skills can help individuals become more conscious of their biases and develop strategies to evaluate information objectively. Individuals can navigate confirmation bias and make more informed decisions by actively seeking diverse perspectives, considering contrary evidence, and engaging in open-minded inquiry.



Confirmation bias has been described as an internal "yes man", echoing back a person's beliefs like Charles Dickens' character Uriah Heep.<sup>i</sup>

### ***Example***

Let's say that Sarah strongly believes in the benefits of a particular diet plan that promotes weight loss. She has been following this diet plan for several months. She has personally experienced positive results, such as losing a few pounds and feeling more energetic.

One day, Sarah comes across an article discussing a scientific study suggesting that her diet plan may not be as

effective as claimed. The study provides evidence and analysis from a reputable source, presenting an alternative perspective on the diet's effectiveness.

However, due to confirmation bias, Sarah unconsciously seeks information confirming her preexisting belief in the diet plan. She may selectively remember success stories from others who have followed the same diet plan or focus on anecdotal evidence supporting its benefits. Sarah might even interpret the study's findings to align with her beliefs, dismissing any contrary evidence or downplaying the study's credibility.

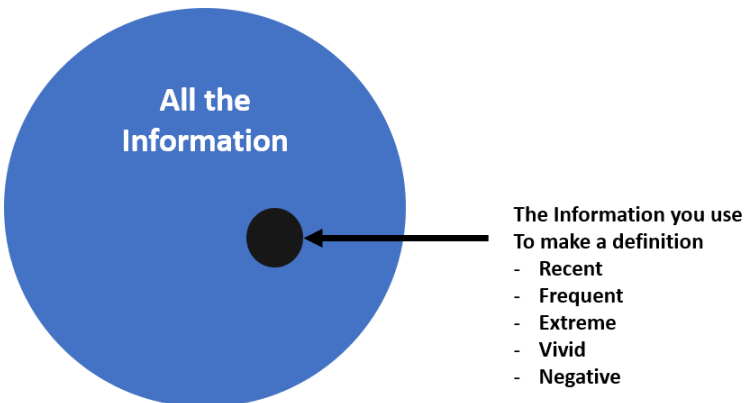
Sarah's confirmation bias leads her to ignore or discount information that challenges her beliefs about the diet plan. Instead, she seeks out and gives more weight to information that supports her preconceived notions. As a result, she maintains her belief in the diet's effectiveness without critically examining the potential limitations or considering alternative viewpoints.

# Availability Heuristic

## Availability bias

The tendency to overestimate the likelihood of events having greater "availability" in memory may be influenced by how recent the memories are or how unusual or emotionally charged they may be.

The availability heuristic, also known as availability bias, is a mental shortcut that relies on immediate examples that come to a person's mind when evaluating a particular topic, concept, method, or decision. It is a cognitive process where individuals judge based on the ease with which relevant examples or instances come to mind.



It is based on the notion that something that can be remembered must be necessary or more important than alternative solutions that cannot be easily recognized. In other words, if the information is readily available in one's memory, it will likely be considered a representative or common occurrence.

As a result, because of the availability heuristic, people tend to bias their judgments heavily toward recent information. This means that new opinions or evaluations are often influenced and skewed by the latest news or events that are more easily accessible in memory.

The availability heuristic can lead to biases in decision-making and judgment, as it may cause individuals to overestimate the likelihood or importance of events or circumstances based solely on their availability in memory. It is essential to be aware of this bias and strive for a more comprehensive and balanced assessment of information and alternatives when making decisions or forming opinions.

### ***Example:***

Imagine you are considering taking a flight to visit a friend in another city. As you start planning, you find a news article about a recent airplane crash. The news story's vivid

details and emotional impact make it easily accessible in your memory.

Due to the availability heuristic, you might overestimate the likelihood of being involved in a plane crash because the news article's recent and emotionally charged memory dominates your thinking. As a result, you might feel hesitant or fearful about taking the flight, even though statistically, air travel is one of the safest modes of transportation.

In this example, the availability heuristic leads to a biased judgment based on the ease with which the negative example of the plane crash comes to mind. It influences your perception of risk and skews your decision-making process, as you give more weight to the recent and emotionally charged memory rather than considering the overall safety record of air travel.

# Anchoring or Focalism

## Anchoring Bias

The anchoring effect refers to the tendency of individuals to rely heavily on a specific piece of information or initial reference point, known as the anchor when making decisions or judgments. This cognitive bias occurs when the initial information presented influences our subsequent thinking and evaluation of a situation.

The term 'anchor' originates from its nautical meaning, where an anchor is a device used to secure a ship and prevent it from drifting. Just as an anchor provides stability to a ship, the anchoring effect provides a reference point that influences our mental stability in decision-making.

Anchors can take various forms, such as numerical values, prices, or opinions. Once an anchor is established, it serves as a mental reference point that influences our subsequent judgments. We adjust our judgments or estimates based on the initial anchor rather than starting from scratch or considering all available information objectively.

The anchoring effect can impact various aspects of decision-making, including negotiations, pricing strategies, and personal judgments. Depending on the initial anchor presented, It can lead to overvaluation and undervaluation of goods, services, or other pieces of information.

Being aware of the anchoring effect can help individuals make more informed decisions by considering multiple sources of information and critically evaluating the initial anchor's relevance and accuracy.

### ***Example***

Imagine you're shopping for a new laptop. You visit a store and come across two models that catch your attention. The first laptop is priced at \$1,500, and the second is priced at \$800.

Now, you're not entirely sure about the actual value of these laptops, but you're looking for a good deal. The \$1,500 laptop serves as an anchor for your decision-making process. Compared to that anchor, the \$800 laptop might seem like a great bargain, even though you don't have any other information about its specifications or market value.



In this scenario, the initial anchor of \$1,500 influences your perception of the \$800 laptop as a more affordable option. The \$800 laptop offers better value for money simply because it's significantly cheaper than the initial anchor.

However, you encountered a different initial anchor, such as a \$500 laptop. In that case, your perception of the \$800 laptop might have been different. The \$800 price tag could seem relatively higher than the new anchor, making you perceive it as less of a bargain.

# Overconfidence Effect

## Egocentric bias

The overconfidence effect is a widely recognized bias where an individual's subjective confidence in their judgments exceeds the factual accuracy of those judgments, mainly when their confidence is relatively high. It is an example of a misperception of subjective probabilities. In the research literature, overconfidence has been defined in three distinct ways:

- Overestimation of one's actual performance: This form of overconfidence involves individuals believing that they will perform better or achieve better outcomes than what their actual abilities or track record indicate. They may overestimate their skills, knowledge, or capabilities in a particular task or domain.
- Overestimation of one's performance relative to others: In this case, individuals overestimate their performance or abilities compared to others. They perceive themselves as more competent, skilled, or superior relative to their peers or competitors, even if objective measures may suggest otherwise.

- Overconfidence in expressing unwarranted certainty about one's beliefs: This form of overconfidence pertains to individuals exhibiting unwarranted certainty or absolute confidence in the accuracy or correctness of their beliefs, opinions, or judgments. They may express strong convictions or unwavering certainty even when the evidence or information supporting their views is limited or ambiguous.

These different manifestations of overconfidence highlight the pervasive nature of this bias and its impact on decision-making, problem-solving, and interpersonal interactions. The overconfidence effect can lead individuals to make overly optimistic predictions, take excessive risks, disregard alternative viewpoints, and make erroneous judgments.

Understanding the overconfidence effect can promote more objective and accurate assessments. By cultivating awareness of one's limitations, seeking diverse perspectives, and fostering a healthy skepticism towards one's judgments, individuals can mitigate the impact of overconfidence and make more informed and balanced decisions.

### ***Example***

Emily, a student preparing for a difficult math exam, exhibits the overconfidence effect in her study predictions. She believes she profoundly understands the material and is highly confident performing well on the exam. Despite knowing that the exam is challenging and requires extensive preparation, Emily feels she has studied enough and is overly optimistic about her performance.

As a result of her overconfidence, Emily decides to allocate less time to studying than she needs. She underestimates the complexity of the exam questions and overestimates her ability to recall information accurately. This leads her to prioritize other activities and engage in fewer study sessions than necessary.

On the exam day, Emily realizes that her overconfidence has misled her. The questions proved to be more difficult than anticipated, and she struggled to answer them correctly. Her initial confidence turns into anxiety and regret as she realizes her overestimation of her knowledge and readiness.

This example demonstrates the overconfidence effect or egocentric bias, wherein individuals overestimate their abilities or knowledge. In Emily's case, her subjective

confidence in her preparation exceeds the accuracy of her understanding of the material. This bias can lead individuals to make suboptimal decisions, such as allocating inadequate time or resources, underestimating challenges, and overestimating their likelihood of success.

Recognizing the overconfidence effect can help individuals adopt a more balanced and realistic perspective. By actively seeking feedback, considering alternative viewpoints, and continually evaluating their performance, individuals can mitigate the negative consequences of overconfidence and make more informed decisions based on accurate self-assessments.

# Halo Effect

## Association fallacy

The halo effect, also known as the halo error, refers to the tendency for positive impressions or evaluations of a person, company, brand, or product in one specific area to influence opinions or feelings about that entity in other unrelated areas. A cognitive bias can lead individuals to form positive or negative judgments based on limited information or a single positive attribute.

The term "halo effect" describes the phenomenon where evaluators or perceivers are influenced by their previous favorable judgments about an individual's performance, personality, or other attributes when forming opinions in unrelated domains.

Edward Thorndike coined the term "halo effect" and highlighted its impact on subjective evaluations and judgments. The halo effect can lead to biased decision-making and judgments based on generalizations and assumptions rather than a comprehensive assessment of the attributes or qualities in question.

The halo effect can prevent individuals from objectively evaluating a person, product, or brand, as they may be swayed by the overall positive impression created by a single positive characteristic. This bias can create unfounded beliefs about the entity's overall quality or desirability, leading to biased perceptions and potentially overlooking any negative aspects or limitations.

Individual preferences, prejudices, ideologies, aspirations, and social perceptions influence the halo effect. It reflects the tendency to use mental shortcuts, known as heuristics, to simplify complex judgments and make quick assessments based on limited information.

To mitigate the impact of the halo effect, it is essential to engage in critical thinking, gather comprehensive information, and evaluate entities based on multiple relevant attributes rather than relying solely on one positive or negative characteristic. By being aware of the halo effect and actively seeking diverse perspectives and information, individuals can make more balanced and informed judgments, leading to more accurate assessments and decisions.

Understanding the halo effect can also help companies and brands' marketing strategies. By carefully managing

their reputation, consistently delivering high-quality products or services, and addressing any negative perceptions or biases, they can minimize the influence of the halo effect and build genuine trust and loyalty among their customers.

### ***Example***

A job interview scenario where an applicant walks into the room: The interviewer notices that the applicant is well-dressed, confident, and has a pleasant smile. These positive attributes create an initial positive impression. As the interview progresses, the interviewer may subconsciously attribute other positive qualities to the applicant, such as intelligence, competence, and professionalism, based solely on the initial positive impression.

As a result of the halo effect, the interviewer may overlook or downplay any shortcomings or weaknesses exhibited by the applicant during the interview. For instance, if the applicant struggles to answer a specific question or lacks experience in a particular area, the positive halo created by the initial impression may lead the interviewer to give the applicant the benefit of the doubt or make excuses for their deficiencies.



In this example, the halo effect can influence the overall evaluation of the applicant's suitability for the job. The positive attributes observed initially overshadow any negative aspects, and the interviewer may be more inclined to view the applicant as highly qualified and capable, even if there is insufficient evidence to support such a conclusion.

This bias can have significant implications, as it may result in hiring a less qualified candidate or overlooking potential red flags. In this case, the halo effect can distort the interviewer's judgment and compromise the fairness and accuracy of the hiring process.

It is important to note that the halo effect can also work in the opposite direction. Suppose the initial impression is negative, such as a poorly dressed or nervous applicant. In that case, the halo effect can lead to an overall negative evaluation, potentially overshadowing the individual's qualifications and capabilities.

The halo effect can influence subjective evaluations and judgments in various domains, including hiring decisions, product evaluations, and social interactions. Recognizing and being mindful of this bias can help individuals make more objective and fair assessments, ensuring that

decisions are based on a comprehensive evaluation of relevant factors rather than being swayed by a single positive or negative attribute.

# Recency Effect

## Memory

Two traditional categories of theories explain the recency effect:

- **Dual-store models** propose that the recency effect occurs because recently presented items are still accessible in a short-term buffer known as short-term storage (STS) in human memory. In dual-store models, the advantage of recent items over earlier items is attributed to their easier retrieval from the STS. Since earlier items have already transitioned to long-term memory, they require more effort to retrieve. According to this view, the recency effect arises from the temporal dynamics of memory retrieval.
- **Single-store models:** Unlike dual-store models, single-store theories suggest that a single mechanism is responsible for the primacy and recency effects in serial position memory tasks. One type of single-store model is based on relative temporal distinctiveness. It posits that the time interval between the study of each list item and the

subsequent test determines the competitive advantage of an item's memory trace during retrieval. In this model, items at the end of the list are assumed to be more distinct and more accessible to recall than those in the middle. Therefore, the recency effect is explained by the enhanced distinctiveness of the most recently presented items.

Both dual-store and single-store models explain the recency effect, emphasizing the temporal accessibility of recently presented items and the relative distinctiveness of items in memory. These theoretical frameworks have contributed to our understanding of memory processes and the factors influencing the retrieval of information from memory.

### ***Example***

Let's say you attend a conference where multiple speakers present their research findings. The conference spans several hours, and each speaker discusses a different topic. As the conference progresses, you listen attentively to each speaker and take notes.

At the end of the conference, you are asked to recall the main points or key findings from each presentation. As you

start recalling the information, you notice that you can easily remember the details of the most recent presentations. Such details occurred towards the end of the conference. The information from these presentations comes to mind quickly and effortlessly.

However, when it comes to remembering the presentations that occurred earlier in the day, you find it more challenging. You have to put in more effort to retrieve the details; some earlier presentations might be more brutal to recall accurately.

This example demonstrates the recency effect. Despite the passage of time and multiple presentations, your memory prioritizes the most recent information. It gives it an advantage in the recall process. The recency effect suggests that information presented towards the end of a sequence or event is more readily accessible and therefore remembered more accurately than earlier information.

# Framing Effect

## Framing effect

The framing effect is a cognitive bias that demonstrates how people's decisions can be influenced by how information is presented to them. Individuals can draw different conclusions or make different choices depending on whether the information is framed positively or negatively.

When information is presented in a positive frame, emphasizing potential gains or benefits, people tend to be more inclined to take risks and pursue options that offer potential rewards. They are motivated to maximize their gains and experience positive outcomes.

On the other hand, when information is presented in a negative frame, focusing on potential losses or drawbacks, individuals tend to be more risk-averse. The fear of incurring losses drives them and are more likely to choose options that minimize risks or avoid potential negative consequences.

The way information is presented can greatly affect the choices people make in different areas such as personal

finance, healthcare, and marketing. When it comes to finances, someone may be more likely to invest in something if they are shown the potential benefits, but may be more hesitant if the focus is on the potential risks. This is known as the framing effect.

In healthcare, the framing of treatment options can influence patients' choices. Presenting a medical procedure as having a higher success rate may make patients more likely to opt for it than when the emphasis is on the potential risks or failures.

Marketers also utilize the framing effect to influence consumer behavior. They may highlight a product's positive features or benefits to encourage purchase decisions, leveraging the tendency to respond more favorably to positive frames.

The framing effect demonstrates the importance of considering how information is presented and how it can shape our decisions. It reminds us that our choices can be influenced not only by the content of the information but also by its framing or presentation. Awareness of this bias can help us make more informed and objective decisions.

### **Example**

Picture yourself contemplating medical treatment for a particular ailment. Your doctor gives you two choices:

**Option A:** This treatment has a 90% success rate. **Option B:** This treatment has a 10% failure rate.

In this case, Option A is framed positively, emphasizing the potential gains or success of the treatment.

**Option B**, on the other hand, is framed negatively, highlighting potential failures or adverse outcomes.

Due to the framing effect, many individuals would be more inclined to choose Option A. The positive frame of a 90% success rate is likely more appealing and reassuring, leading them to believe that this treatment is highly likely to work for them.

However, if the information is reframed, the decision might change:

**Option C:** This treatment has a 10% failure rate. **Option D:** This treatment has a 90% success rate.

Although the information is the same as before, the framing is reversed. Now, Option C is presented positively,



emphasizing a low failure rate. In contrast, Option D is presented negatively, focusing on the potential for failure.

In this scenario, individuals may be more likely to choose Option C, perceiving it as a safer choice with a lower risk of failure.

# Sunk Cost Fallacy

## Sunk cost fallacy

In economics and business decision-making, sunk costs (also referred to as retrospective costs) are already incurred and cannot be recovered. These costs contrast to prospective costs, which represent future expenses that can be avoided or mitigated through specific actions. Sunk costs are the amounts of money, time, or resources spent in the past and are now irretrievable.

From an economic standpoint, sunk costs are considered irrelevant to rational decision-making regarding future actions. This is because sunk costs are irreversible and cannot be changed, regardless of the choices made moving forward. Rational decision-making focuses on prospective costs and benefits, analyzing the potential gains or losses associated with future decisions without considering past expenditures.

However, in everyday life, people often find it challenging to ignore sunk costs when making decisions. For example, when considering whether to repair a car or a house, individuals may factor in the money they have previously

invested in these assets. This inclusion of sunk costs in decision-making is a deviation from economic rationality. Still, it can be attributed to various psychological biases, such as loss aversion and the desire to avoid feelings of regret.

Understanding the concept of sunk costs is crucial in business and economic analysis. By recognizing that sunk costs should not be a primary factor in decision-making, individuals and organizations can make more rational choices and allocate resources effectively based on prospective costs and benefits.



*The sunk cost fallacy has also been called the "Concorde fallacy": the British and French governments took their past expenses on the costly supersonic jet as a rationale for continuing the project, as opposed to "cutting their losses".<sup>ii</sup>*

### **Example**

Suppose you've made the decision to open a small café. To get started, you invest a considerable amount of money in renovating the space, acquiring equipment, and stocking up on inventory. However, after a few months of operation, you realize that the café is not generating enough revenue to cover the expenses and turn a profit.

At this point, you are faced with a decision: continue operating the café and hope for a turnaround, or cut your losses and close the business.

If you were to solely consider the sunk costs—the money you have already invested—you might be inclined to keep the café open, hoping it will eventually become profitable and you can recoup your initial investment. This decision would be influenced by the emotional attachment to the money and effort you have already put into the business.

However, from a rational economic standpoint, sunk costs should not be the primary factor in your decision-making. Instead, it would help if you focused on the future costs and benefits. You would need to assess the current and future potential revenues, expenses, market conditions, and other relevant factors to determine whether it is financially viable to continue operating the café.

You can make a more rational decision by recognizing that sunk costs are irrelevant to the business's prospects. Suppose the prospective costs outweigh the potential benefits, and there is little likelihood of turning the business around. In that case, it may be wise to cut your losses and close the café, even though you have already invested a significant amount of money.

# Hindsight Bias

## Hindsight bias

Hindsight bias, also known as the "knew-it-all" phenomenon or creeping determinism, refers to the common tendency of people to view past events as more predictable than they were. It is the belief that after an event has occurred, individuals would have foreseen or known its outcome, often with a high level of certainty. This bias can lead to a distorted memory of what was known or believed before the event. Hindsight bias significantly contributes to overconfidence in people's ability to predict future events.

Examples of hindsight bias can be observed in various contexts. When describing the outcome of battles, historians may unconsciously attribute more predictability to the events based on the knowledge of the eventual outcome. Similarly, physicians may recall clinical trials and mistakenly believe they could have accurately predicted the results, even if the outcome was uncertain. In the justice system, individuals may assign responsibility for accidents based on the assumption that the event was foreseeable, influenced by their hindsight bias.

The hindsight bias can have profound implications. It can lead individuals to overestimate their foresight and decision-making abilities, affecting their judgment and decision-making in future situations. Recognizing and understanding this bias is crucial for mitigating its effects and promoting more accurate assessments of past events and future predictions.

### ***Example***

Sarah is watching a soccer match between Team A and Team B. Team A starts to dominate as the game progresses, scoring several goals and displaying superior skills. Toward the end of the match, Sarah remarks to her friend, "I knew all along that Team A would win. They were the stronger team."

When the match began, Sarah had no way of accurately predicting the outcome. However, after seeing Team A's dominant performance and the final result, Sarah's memory of her initial thoughts becomes distorted by hindsight bias. She now believes she possessed foresight and always knew that Team A would be victorious.

This example highlights how hindsight bias can lead individuals to revise their beliefs about past events, making them appear more predictable or expected than

they were. It showcases the tendency to overestimate one's ability to foresee outcomes after they have occurred, which can influence perceptions of personal knowledge and decision-making abilities.



# Loss Aversion

## Prospect theory

Loss aversion is a cognitive bias that describes the tendency of individuals to strongly prefer avoiding losses over acquiring equivalent gains. It suggests that the psychological impact of losing something is more significant than the pleasure of gaining something of equal value. Loss aversion is a fundamental principle in behavioral economics and has important implications for decision-making.

To understand loss aversion, consider the following example: Imagine you have the opportunity to participate in a gamble. In one scenario, you stand to win \$100; in another, you stand to lose \$100. According to loss aversion, individuals would typically experience the emotional pain of losing \$100 more intensely than the pleasure of gaining \$100. Consequently, they may avoid the gamble altogether or require a higher potential gain to justify taking the risk.

Loss aversion differs from risk aversion, the general aversion to taking risks regardless of potential gains or

losses. Loss aversion focuses explicitly on the asymmetry between losses and gains, highlighting that losses substantially impact individuals' well-being and decision-making.

Research conducted by psychologists Amos Tversky and Daniel Kahneman, pioneers in the field of behavioral economics, played a significant role in introducing and studying loss aversion. Their seminal work on prospect theory in the 1970s shed light on the influence of emotions and framing effects on decision-making, emphasizing the pronounced effect of loss aversion on people's choices.

Empirical studies have found that the negative emotional impact of losses can be approximately twice as powerful as the positive emotional impact of equivalent gains. This asymmetry in the perception of losses and gains has been observed in various contexts, including financial decision-making, consumer behavior, and investment choices.

Loss aversion has significant implications for understanding human behavior in economic and everyday life. It helps explain why individuals may be reluctant to take risks, holds onto underperforming investments or possessions, and exhibit a bias towards maintaining the

status quo rather than pursuing potentially beneficial changes.

By recognizing the influence of loss aversion, individuals and policymakers can make more informed decisions and consider the emotional factors that shape their choices. Understanding the power of loss aversion can help mitigate potential biases and lead to more rational and effective decision-making.

### ***Example***

Imagine you have a prized possession—an antique watch—that you are considering selling. You bought the watch for \$500, but its market value is uncertain. A potential buyer offers you \$400 for the watch.

Loss aversion suggests that you may be hesitant to sell the watch at this price because the potential loss of \$100 feels more significant than the gain of \$100. Even though the objective gain and loss are equal in monetary terms, the emotional impact of losing \$100 outweighs the pleasure of gaining the same amount.

In this scenario, loss aversion may lead you to resist selling the watch unless you receive a higher offer. You might be more inclined to hold onto the watch, hoping for a better

deal or to avoid the regret and emotional discomfort associated with perceiving a loss.

This example demonstrates how loss aversion can influence decision-making by making individuals more cautious and risk-averse when faced with potential losses. By recognizing this bias, individuals can become more aware of their tendency to avoid losses and make more objective and rational choices.

Remember, loss aversion can manifest in various situations beyond financial decisions, such as career choices, personal relationships, and even everyday choices like selecting a menu item. Its impact on decision-making highlights the importance of understanding our biases to make more informed and balanced judgments.

# Gambler's Fallacy

## Logical fallacy

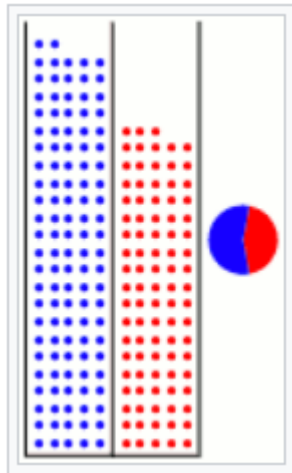
The gambler's fallacy, also known as the Monte Carlo fallacy or the maturity of odds fallacy, refers to the tendency to believe that future probabilities are modified by past events when, in reality, they remain unchanged.

This cognitive bias involves the false belief that a specific event, which has occurred more frequently than usual in the past, is less likely to occur in the future (or vice versa), even though it has been established that the probability of such events does not depend on past events. Events that exhibit this property of historical independence are referred to as statistically independent.

The gambler's fallacy is often associated with games of chance or gambling scenarios. For example, someone may believe that the next throw of the dice is more likely to result in a six because there have been fewer than the usual number of sixes in recent rolls. This belief disregards that each dice roll is an independent event with the same probability of landing on any specific number. The fallacy arises from a misunderstanding of probability and a

tendency to seek patterns or trends in random processes. People may mistakenly assume that past outcomes influence future outcomes, leading to misguided predictions or expectations.

Understanding the gambler's fallacy is essential for making informed decisions and avoiding irrational beliefs about probability. Recognizing that past occurrences do not influence independent events helps individuals make more accurate risk assessments and avoid potential pitfalls in domains where chance and probability play a role.



*Over time, the proportion of red/blue coin tosses approaches 50-50, but the difference does not systematically decrease to zero<sup>iii</sup>*

### ***Example***

Let's say you're at a casino playing roulette, and the last five spins of the roulette wheel have resulted in red numbers. Observing this streak of red numbers, you start believing that a black number is more likely to appear on the next spin. You think, "The wheel is due for a black number since there have been so many reds in a row."

However, the reality is that each spin of the roulette wheel is an independent event, and the probability of landing on either a red or black number remains the same for each spin. The previous outcomes of red numbers do not influence the probability of the next spin. By falling for the gambler's fallacy, you may place a larger bet on black, assuming it will likely come up next. Unfortunately, this belief is unfounded, as the odds of landing on a red or black number are always the same on each wheel spin. The gambler's fallacy can lead to poor decision-making in gambling and other situations involving probability. It's essential to recognize that past events do not alter the probabilities of future events in independent systems, and each event should be evaluated based on its probability rather than influenced by previous outcomes

# Self-Serving Bias

## Attribution bias

The self-serving bias refers to individuals taking more credit for their successes than their failures, a typical cognitive or perceptual process. It also manifests in their tendency to interpret ambiguous information to benefit their interests. This bias is closely related to the group-serving bias, which refers to the tendency to favor one's group over others.

Self-serving bias is the tendency to view oneself positively, fueled by the desire to maintain and boost self-esteem. This can lead people to have an excessively positive perception of themselves. It involves attributing successes to one's abilities, efforts, and failures to external factors beyond control. Individuals protect their self-esteem from threats and harm by rejecting negative feedback, emphasizing strengths and accomplishments, and downplaying failures and mistakes. These cognitive and perceptual tendencies create and perpetuate illusions and misconceptions.



Humans have a fundamental need for self-esteem, and the self-serving bias helps them maintain positive self-perceptions. However, it can also distort reality and hinder accurate self-evaluation and growth. Recognizing and mitigating the influence of self-serving bias can promote more objective assessments and a clearer understanding of one's strengths and weaknesses.

### ***Example***

Let's picture a sales representative named Alex, who is employed by a business that markets a new item. Alex has been assigned a sales target for the month, and they are determined to achieve it. After a month of hard work and persuasive pitches, Alex successfully surpasses the sales target and receives recognition from their manager and colleagues.

When asked about their success, Alex attributes it to their exceptional sales skills, product knowledge, and relentless efforts. They believe their persuasive abilities and dedication were vital to outstanding sales performance. This attribution reinforces their positive self-image and boosts their self-esteem.

However, Alex still needs to meet the sales target the following month. Despite the setback, they attribute the

poor performance to external factors such as a challenging market, increased competition, or economic conditions. They downplay any personal responsibility and maintain that their sales skills and efforts are still practical, despite unfavorable results.

This bias can be seen in various domains, such as academics, sports, or personal achievements, where individuals internalize successes and externalize failures. It allows individuals to maintain a positive self-image and preserve their self-esteem. Still, it can also lead to distorted perceptions of reality and hinder personal growth and improvement.

# Dunning-Kruger Effect

## Dunning-Kruger effect

The Dunning-Kruger effect refers to a cognitive bias characterized by two distinct tendencies: the overestimation of abilities by individuals with low skill levels and the underestimation of abilities by experts in a given domain. This phenomenon highlights the inherent challenges in accurately assessing one's competence.

The effect was named after psychologists David Dunning and Justin Kruger, who conducted pioneering research on this bias. They found that individuals with limited skills or knowledge often exhibit an apparent overconfidence in their abilities. Due to their lack of expertise, they cannot recognize their deficiencies and thus mistakenly perceive themselves as highly competent.

Conversely, individuals with high competence or expertise tend to underestimate their abilities. This occurs because experts possess a deeper understanding of the complexity and nuances within a domain, making them more aware of their limitations and the vast amount of knowledge yet to be acquired. As a result, they may underestimate their

abilities relative to others or fail to recognize their exceptional competence.

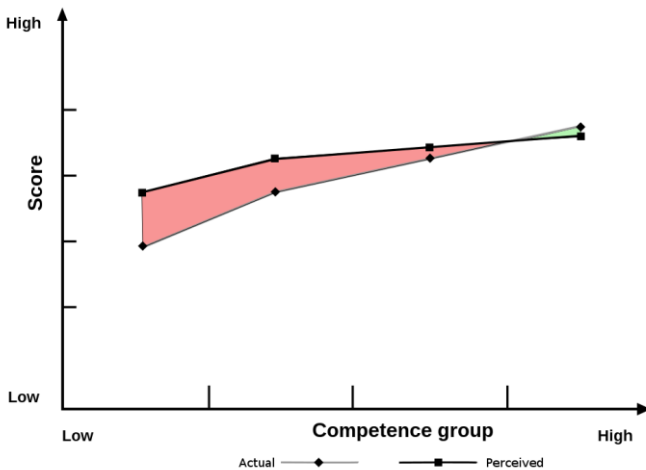
The Dunning-Kruger effect is commonly studied by comparing individuals' performance self-assessments with objective measures. For example, participants may complete a task or assessment and subsequently rate their performance. These self-assessments are then compared to actual performance or evaluated against established benchmarks to determine the degree of bias present.

This effect has implications across various fields and endeavors. In academic settings, students with limited understanding may overestimate their knowledge, leading to challenges in learning and growth. In professional contexts, less skilled individuals may exhibit unwarranted confidence, potentially impacting decision-making, problem-solving, and overall performance. Meanwhile, experts may exhibit modesty or self-doubt, despite their high level of competence, which could hinder their advancement or influence their willingness to take on new challenges.

Understanding the Dunning-Kruger effect helps promote humility, self-awareness, and a more accurate assessment of one's abilities. It highlights the importance of seeking

feedback, ongoing learning, and recognizing the limits of one's expertise. By embracing a growth mindset and acknowledging the complexity of skills and knowledge, individuals can strive for continuous improvement and make more informed judgments about their abilities and achievements.

**Dunning–Kruger Effect**



Relation between average self-perceived performance and average actual performance on a college exam. The red area shows the tendency of low performers to overestimate their abilities. Nevertheless, low performers' self-assessment is lower than that of high performers.<sup>iv</sup>

### ***Example***

Meet Mark, an aspiring guitarist who has recently picked up the instrument. Mark has been practicing diligently for a few weeks and believes he is already a talented guitarist. He confidently showcases his skills to friends and family, often boasting about his ability to play complex melodies and solos.

Mark's guitar skills are limited, and his performances could be more impressive. However, due to his limited experience and lack of awareness about the depth of skill required to master the instrument, he overestimates his abilities. Mark falls victim to the Dunning-Kruger effect, as he lacks the knowledge and expertise to assess his guitar-playing proficiency accurately.

On the other hand, we have Emily, an experienced and skilled guitarist who has played for many years. Despite her impressive abilities and extensive knowledge of music theory, Emily often doubts her skills. She underestimates her talent, comparing herself unfavorably to other accomplished guitarists and believing she has much more to learn.

In this example, Mark represents the unskilled individual who overestimates his abilities. In contrast, Emily

represents the expert who underestimates her abilities. The Dunning-Kruger effect highlights the stark contrast in self-assessments between individuals with differing levels of skill and expertise.

By understanding the Dunning-Kruger effect, individuals like Mark can become more aware of their limitations and strive for genuine growth and improvement. They can seek guidance from more experienced individuals, engage in deliberate practice, and gradually better understand their abilities. Similarly, individuals like Emily can recognize their accomplishments, acknowledge their expertise, and develop a more confident outlook.

The Dunning-Kruger effect serves as a reminder that self-assessment should be approached with humility and a willingness to learn. It emphasizes the importance of seeking feedback, being open to constructive criticism, and continuously expanding one's knowledge and skills. By doing so, individuals can strive for a more accurate understanding of their abilities and make progress toward mastery in their respective domains.

# Social-Desirability Bias

## Social-desirability bias

In social science research, social desirability bias is a specific type of response bias characterized by the tendency of survey respondents to answer questions in a manner that they believe will be viewed favorably by others. It manifests as a tendency to overstate positive or socially desirable behavior while understating or underreporting harmful or socially undesirable behavior. This bias poses a significant challenge when conducting self-report surveys.

When individuals are aware that others will evaluate their responses, they may be motivated to present themselves in a more favorable light, adhering to societal norms or expectations. This can result in respondents providing answers that they perceive as socially desirable, even if those responses do not accurately reflect their true thoughts, feelings, or behaviors.

The impact of social desirability bias extends beyond individual survey responses. It can affect the interpretation of data related to average tendencies and



individual differences. Suppose a significant number of respondents exhibit a social desirability bias. In that case, it can lead to an inflated perception of positive behaviors and a potential underestimation of harmful or undesirable behaviors within a population.

Researchers employ various strategies to mitigate the influence of social desirability bias, such as using indirect questioning techniques, ensuring anonymity or confidentiality, and employing validation measures. These approaches aim to create a more neutral and non-judgmental survey environment, encouraging respondents to provide more honest and accurate responses.

Awareness of social desirability bias is crucial for researchers and survey designers to interpret and analyze data accurately. By acknowledging the potential impact of this bias and implementing appropriate methods to mitigate its effects, researchers can strive for a clearer understanding of individuals' attitudes, behaviors, and experiences in social science research.

### ***Example***

In a study examining health behaviors and habits, participants were asked to self-report their weekly

exercise frequency. Due to social desirability bias, some respondents may feel inclined to present themselves as more physically active than they are, believing that regular exercise is socially desirable and viewed positively by others.

For instance, Sarah, one of the study participants, may engage in exercise only once or twice a week but feels pressured to provide a response that aligns with societal expectations. As a result, she overstates her exercise frequency by reporting that she exercises five times a week, which may not accurately reflect her actual behavior.

Social desirability bias in this scenario can distort the data collected from self-report surveys. Suppose a significant number of participants exhibit this bias. In that case, it may lead to overestimating exercise frequency within the studied population. Consequently, the interpretation of the average tendency for exercise and individual differences may be skewed, potentially leading to inaccurate conclusions or recommendations regarding health behaviors.

To mitigate the impact of social desirability bias, researchers can implement strategies such as ensuring

participant anonymity, emphasizing confidentiality, or utilizing indirect questioning techniques. By creating an environment that encourages honest and genuine responses, researchers can minimize the potential distortion caused by social desirability bias and obtain more accurate data on participants' exercise habits.

Recognizing and accounting for social desirability bias is essential to obtain reliable and valid findings in social science research. By employing appropriate measures to address this bias, researchers can better understand individuals' proper behaviors, attitudes, and experiences, ultimately enhancing the quality and validity of their research outcomes.

# Illusory Correlation

## Apophenia

In psychology, spurious correlation refers to perceiving a relationship between variables (usually people, events, or behaviors) even though no such relationship exists. It is a cognitive bias where individuals mistakenly attribute a causal or meaningful connection between two or more variables when, in reality, the correlation is coincidental or non-existent.

One reason for spurious correlations is the tendency for rare or novel occurrences to capture people's attention. When a rare event coincides with another variable, individuals may erroneously perceive a meaningful relationship between the two, even though it is merely a chance occurrence.

Spurious correlations play a role in the formation and maintenance of stereotypes. A study conducted by Hamilton & Rose (1980) found that stereotypes can lead individuals to expect certain groups and characteristics to be associated and subsequently overestimate the frequency with which these correlations occur. These

stereotypes can persist even without direct contact between the holder of the stereotype and the group being stereotyped.

Recognizing and challenging spurious correlations is essential to avoid making incorrect assumptions or generalizations based on coincidental associations. By understanding the potential for spurious correlations, individuals can better understand the proper relationships between variables and avoid perpetuating stereotypes or false beliefs.

### ***Example***

Suppose a study explores the correlation between ice cream sales and crime rates in a particular city. The data collected shows that during the summer months, when ice cream sales are high, crime rates also tend to be high. Based on this correlation, one might mistakenly conclude that ice cream consumption somehow causes an increase in criminal activity.

However, this correlation is spurious. The underlying factors behind this correlation are the summer season and increased outdoor activity. During the summer, more people are out and about, leading to higher ice cream sales and increased opportunities for criminal activity. The

correlation between ice cream sales and crime rates is coincidental and not causally linked.

In this example, it is crucial to recognize that there is no direct relationship between ice cream consumption and criminal behavior. The spurious correlation arises due to the confounding variable of the summer season. Failing to understand this spurious correlation could lead to misguided policies or assumptions about the causes of crime.

# Mere-Exposure Effect

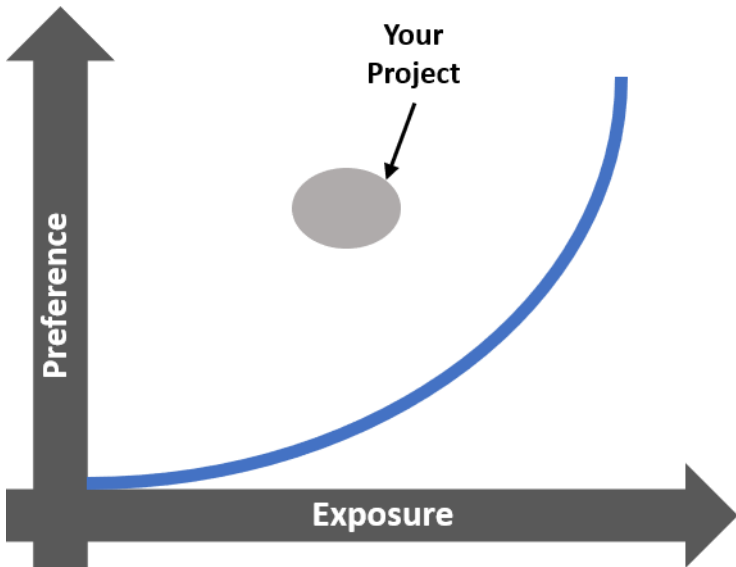
## Familiarity principle

The mere-exposure effect, also known as the familiarity principle, refers to the tendency to express undue liking for things merely because of familiarity. It is a psychological phenomenon where individuals develop a preference for objects, people, or stimuli simply because they have been exposed to them repeatedly.

Extensive research in various domains of psychology has examined this phenomenon. Studies have shown that individuals tend to favor stimuli, such as words, Chinese characters, paintings, pictures of faces, geometric figures, and even sounds, that they have encountered multiple times. The more exposure they have to these stimuli, the more positively they evaluate and appreciate them.

The mere-exposure effect plays a significant role in interpersonal attraction. Research has found that the more we encounter someone, the more we tend to like and find them attractive. This can impact how we interact with others and build relationships since we often develop positive feelings toward those we know.

It is important to note that the mere-exposure effect operates subconsciously, meaning that individuals may not be aware of its influence on their preferences and judgments. The effect highlights the role of familiarity in shaping our perceptions and preferences, even when objective qualities or characteristics are not considered.



Understanding the mere-exposure effect can help us recognize and critically evaluate our preferences, ensuring that we do not overly favor or judge things solely based on familiarity. It reminds us to approach new experiences with an open mind and consider other factors beyond



mere exposure when forming opinions or making decisions.

### ***Example***

Let's say there is a new song that a famous artist has just released. When people first hear the song, they might not firmly believe it. However, as the song gains more airplay on the radio and becomes more widely heard, individuals are exposed to it repeatedly.

Over time, due to the mere-exposure effect, people may start to develop a liking for the song. They might catch themselves humming along to the tune or finding it catchy, even if they didn't initially react strongly to it. Repeated exposure to the song creates a sense of familiarity, contributing to increased preference for the song.

In this example, the mere-exposure effect demonstrates how people's liking for the song can be influenced by their light exposure to it. It showcases that our preferences and attitudes can be shaped by being repeatedly exposed to something, regardless of the song's objective quality or original appeal.

The influence of familiarity is not restricted to music alone; it also applies to other areas, such as advertising, product

marketing, and social interactions. The more we are exposed to something, the more inclined we become toward it, showcasing the significant impact of familiarity on our perceptions and preferences.

# Choice-Supportive Bias

## Confirmation bias

Choice-supportive bias, also known as post-purchase rationalization, refers to the tendency of individuals to retroactively attribute favorable qualities to a chosen option while devaluing the options they didn't choose. This cognitive bias occurs after a decision has been made and can influence how people perceive and remember their choices.

For instance, let's consider a person who chooses option A over option B. After making the decision, they may downplay any shortcomings or failures associated with option A, focusing instead on its positive aspects. At the same time, they might exaggerate or amplify the flaws of option B, attributing new shortcomings to it that were not initially considered.

This bias can be attributed to the desire for consistency and the need to justify one's decisions. By enhancing the positive attributes of the chosen option and devaluing the alternatives, individuals can reduce any cognitive dissonance that arises from their decision.

The memory of a decision plays a significant role in how individuals feel about their choices, including the level of satisfaction or regret they experience. Unfortunately, research indicates that decision-making and subsequent memory formation are susceptible to predictable biases.

Choice-supportive bias can influence an individual's perception of their decision and future decision-making processes. By reinforcing the positive aspects of their chosen option, people may become more likely to choose it again in similar situations, even if objectively better alternatives exist.

Awareness of choice-supportive bias can help individuals approach decision-making with greater objectivity. By actively considering the strengths and weaknesses of different options, individuals can strive for more balanced evaluations and reduce the impact of this bias on their decision outcomes.

### ***Example***

Suppose Sarah is shopping for a new smartphone. After careful consideration, she chooses Option A, a famous brand known for its sleek design and user-friendly interface, over Option B, a competitor with similar features.

Following her purchase, Sarah notices minor flaws with Option A, such as occasional lagging and shorter battery life. However, she downplays these issues and focuses on the positive aspects she initially considered, such as the brand reputation and aesthetics.

On the other hand, Sarah starts to nitpick the flaws of Option B, emphasizing its bulkier design and fewer pre-installed apps. She attributes new shortcomings to Option B, such as lower durability and limited software updates.

Despite both options' objective strengths and weaknesses, Sarah's memory and perception of her decision become biased. She retrospectively attributes even more positive qualities to Option A, conveniently forgetting or downplaying its flaws. At the same time, she magnifies the negative aspects of Option B, reinforcing her belief that she made the right choice.

This choice-supportive bias allows Sarah to maintain consistency in her decision and reduce cognitive dissonance. It also influences her future smartphone choices, as she may be more inclined to choose Option A again, even if other options offer better features or improved performance.

Choice-supportive bias can distort one's perception and memory of a decision. By selectively emphasizing the positive aspects of the chosen option and devaluing the alternatives, individuals reinforce their belief in the correctness of their choice and validate their decision-making process.

# Bandwagon Effect

## Conformity bias

The bandwagon effect refers to the inclination of individuals to adopt certain behaviors, styles, or attitudes simply because they observe others doing so. It represents a cognitive bias in which the spread of specific actions and beliefs can influence public opinion or behavior. This psychological phenomenon highlights how thoughts, ideas, fads, and trends gain momentum and popularity as more people engage in them. Essentially, the bandwagon effect occurs when individuals "jump on the bandwagon" and adopt certain beliefs or behaviors without critically evaluating the underlying evidence.

The bandwagon effect is rooted in the human tendency to seek social validation and conform to group norms. People often look to others for cues on how to behave or what to believe, especially when they perceive a significant number of others engaging in a particular behavior or holding a specific belief. This can lead to a domino effect, where the adoption of certain behaviors or attitudes spreads rapidly through social networks.

It is important to note that the bandwagon effect does not necessarily imply that the adopted behaviors or beliefs are inherently valid or based on sound evidence. Instead, the influence of the crowd and the desire for social acceptance often play a significant role in shaping individual choices and actions.

### ***Example***

Picture a gathering of friends discussing a popular diet trend that has gained a lot of attention on social media. The diet promises exceptional health benefits and weight loss outcomes, with many influencers and celebrities advocating for it. As the group hears about the diet's popularity and sees others around them adopting it, they feel compelled to join in, despite having limited knowledge about its effectiveness or scientific basis.

In this scenario, the follower effect is at play as the group of friends adopts the diet trend primarily because they observe others doing so and perceive it as a popular and socially acceptable choice. The increasing number of people endorsing and practicing the diet creates a sense of social validation, leading friends to jump on the bandwagon without critically evaluating the underlying evidence or considering individual dietary needs.



By succumbing to the follower effect, the friends may overlook the importance of personalized nutrition, scientific research, and professional advice when making dietary choices. Instead, they prioritize conformity and social acceptance, potentially compromising their health and well-being.

This example demonstrates how the follower effect can influence individuals' decisions and behaviors, leading them to adopt certain practices or beliefs simply because they observe others doing the same. It highlights the influence of social influence and the desire for acceptance in shaping individual choices, even if they may not be supported by robust evidence or align with personal needs and preferences.

Recognizing the bandwagon effect can help individuals maintain critical thinking, consider diverse perspectives, and make informed decisions based on their values, knowledge, and evaluation of the available evidence. It encourages individuals to question prevailing trends and seek reliable information before embracing new behaviors or beliefs solely based on popularity or social endorsement.

# Negativity Bias

## Memory

Negativity bias, also known as the negativity effect, refers to a psychological phenomenon in which individuals tend to recall and be affected by unpleasant memories more strongly than positive ones. It suggests that negative experiences or stimuli significantly impact an individual's psychological state and cognitive processes more than neutral or positive experiences of equal intensity.

The bias towards negativity implies that negative thoughts, emotions, or social interactions leave a more lasting impression and substantially influence an individual's behavior and perceptions. For instance, a single adverse event or criticism may significantly impact someone's mood and self-esteem more than multiple positive experiences or compliments.

Research has shown that negativity bias operates in various domains, including forming impressions and evaluations, attentional processes, learning and memory, and decision-making. It is believed to have evolutionary

roots, as it served as a survival mechanism to prioritize potential threats and dangers in our ancestral past.

In everyday life, the negativity bias can be observed in how individuals tend to dwell on adverse events, anticipate potential risks and losses more strongly than potential gains, and have a stronger memory for negative or traumatic experiences. Understanding this bias can provide insights into human cognition, behavior, and how we process and remember information in different contexts.

### ***Example***

Picture yourself at a social event where you come across many new faces. During the course of the evening, you have numerous conversations and exchanges. Unfortunately, one of these interactions didn't go well. The individual you conversed with was impolite and uninterested.

Despite having many positive and pleasant conversations throughout the event, the memory of that negative encounter will likely stand out more prominently in your mind. You may find yourself replaying the details of that unpleasant interaction, recalling the tone of voice, and feeling the negative emotions associated with it.

On the other hand, you may struggle to recall specific details or emotions from the other positive interactions you had during the gathering. Although those positive experiences may have been enjoyable, they may not have left as strong of an impression as the negative ones.

The negative encounter carries more weight in your memory and influences your overall perception of the event. It highlights how our minds are more prone to dwell on and remember negative experiences, even when positive experiences outnumber them.

It's worth mentioning that the negativity bias doesn't imply that we are entirely negative or that positive experiences are insignificant. Instead, it indicates that negative experiences have a more robust and longer-lasting effect on our thoughts, emotions, and memories.

# Fundamental Attribution Error

## Attribution bias

In the field of social psychology, there is a phenomenon called fundamental attribution error (FAE). This occurs when people tend to underestimate the impact of situational and environmental factors on a person's behavior while giving too much weight to their personality and disposition as the cause. It is also referred to as correspondence bias or attribution effect. This effect has been described as 'the tendency to believe that what people do reflects who they are,' that is, to over-attribute their behavior (what they do or say) to their personality and to subordinate it to the situation or context.

The mistake is to view a person's actions solely as an expression of their character rather than viewing them to some extent as an expression of their personality and explaining them mainly in terms of circumstances. It is a circular reasoning in which the answer to the question 'Why would he do that?' is 'Because he would.'

The fundamental attribution error leads the observer to make dispositional attributions, focusing on Sarah's

personality and character while neglecting the situational influences that may have contributed to her actions.

By understanding the fundamental attribution error, we can become more aware of our tendency to overvalue personality explanations and underappreciate situational factors when interpreting others' behavior. It reminds us to consider the context and circumstances in which behaviors occur, leading to a more accurate understanding of people's actions and avoiding hasty judgments based solely on dispositional attributions.

### ***Example***

Imagine driving in heavy traffic, and suddenly another driver cuts you off and aggressively honks their horn at you. Your immediate reaction might be attributing their behavior to their personality, thinking they are rude, aggressive, or inconsiderate.

However, you may not consider the situational factors that could have influenced their behavior. Perhaps the driver ran late for a necessary appointment or received an urgent phone call. Maybe they were distracted by a crying child in the car or dealing with a personal emergency. These situational factors could have contributed to their

aggressive driving rather than solely being a reflection of their character.

The fundamental attribution error would lead you to overvalue dispositional attributions, focusing on the driver's personality traits while overlooking the potential impact of situational factors.

By recognizing this bias, you can remind yourself to consider the broader context and factors that may have influenced someone's behavior before judging their character. It helps to avoid jumping to conclusions based solely on dispositional attributions. It encourages a more nuanced understanding of human behavior.

# ALGORITHMIC BIAS

A brief introduction to algorithmic bias as it is becoming increasingly important.

## **What is Algorithmic Bias?**

Algorithmic bias describes systematic and repeatable errors in a computer system that lead to unfair results, favoring one arbitrary group of users over others. Bias can arise from many factors, including but not limited to algorithm design or unintended or unanticipated use or decisions regarding how data are coded, collected, selected, or used to train the algorithm. For example, algorithmic biases have been observed in search engine results and social media platforms. These biases can have effects ranging from unintentional privacy violations to reinforcing social biases related to race, gender, sexuality, and ethnicity. However, the study of algorithmic bias focuses primarily on algorithms that reflect "systematic and unfair" discrimination. Moreover, this bias has been addressed in legal frameworks such as the European Union General Data Protection Regulation (2018)<sup>v</sup> and the proposed Artificial Intelligence Act (2021)<sup>vi</sup>.



As algorithms expand their ability to organize society, politics, institutions, and behavior, sociologists have become concerned with how data's unpredictable output and manipulation can affect the physical world. Because algorithms are often viewed as neutral and unbiased, they can falsely purport greater authority than human expertise (in part due to the psychological phenomenon of automation bias). In some cases, reliance on algorithms can substitute for human accountability for their outcomes. In addition, bias can enter algorithmic systems due to preexisting cultural, social, or institutional expectations, technical limitations of their design, or through use in unanticipated contexts or by target audiences not considered in the original design of the software.

Algorithmic biases have been cited in cases ranging from election results to the spread of hate speech online. They have also occurred in criminal justice, health care, and hiring, reinforcing existing racial, socioeconomic, and gender biases. For example, the relative inability of facial recognition technology to accurately identify dark-skinned faces has been linked to numerous wrongful arrests of black males, a problem attributed to unbalanced data sets. Difficulties in understanding, exploring, and detecting algorithmic bias exist due to the proprietary nature of

algorithms, which are typically treated as trade secrets. Even when full transparency is provided, the complexity of specific algorithms presents a barrier to understanding how they work. In addition, algorithms may change or respond to inputs or outputs in ways that are not predictable or easily reproducible for analysis. In many cases, even within a single website or application, there is no single "algorithm" to study but rather a network of many related programs and data inputs, even between users of the same service.

### ***Types of Algorithmic Bias***

#### **Technical**

Technical bias emerges from various limitations inherent in programs, computing power, system design, or other technological factors. For instance, consider a scenario where a search engine presents three results per screen. In such cases, the first three results may receive a slightly greater emphasis compared to the remaining three, as observed in the context of airline price displays. Another illustration involves software that relies on random number generation to ensure a fair distribution of outcomes. However, the mechanism for generating random numbers lacks true randomness. In that case, it

can introduce bias, potentially favoring items positioned at the beginning or end of a list.

### Correlations

When large data sets are compared, unpredictable correlations can arise. For example, data collected on Internet browsing behavior may match signals that flag sensitive data (such as race or sexual orientation). By selecting specific behaviors or browsing patterns, the result would be almost identical to discrimination by using direct race or sexual orientation data. In other cases, the algorithm concludes correlations without understanding those correlations. For example, a triage program gave asthmatics with pneumonia a lower priority than asthmatics without pneumonia. The program algorithm did this because it simply compared survival rates: asthmatics with pneumonia have the highest risk. For the same reason, asthmatics in hospitals usually receive the best and most immediate treatment.

### Pre-existing

Pre-existing bias in an algorithm is a consequence of underlying social and institutional ideologies. Such ideas can influence or create personal biases in individual designers or programmers. Poorly selected input data or simply data from a biased source will affect the results

produced by machines. Coding pre-existing biases into the software can preserve social and institutional biases that, without correction, could be repeated in all future applications of the algorithm.

### Emergent

Emergent biases result from using and relying on algorithms in new or unexpected contexts. Algorithms may not have been adapted to account for new forms of knowledge, such as new drugs or medical breakthroughs, new laws, business models, or changing cultural norms. This can result in groups being excluded by the technology, with no clear indication of who is responsible for their exclusion. Similarly, problems can arise when training data (the samples "fed" to a machine, which it uses to model certain conclusions) do not match the contexts an algorithm encounters in the real world.

### Unexpected use

When unexpected audiences use an algorithm, bias can occur. For example, machines may assume that users can read, write, or understand numbers or that they identify with an interface through metaphors they do not understand. These exclusions can be exacerbated as biased or exclusionary technologies become more deeply integrated into society.

## Feedback loops

Emergent biases can also lead to a feedback loop or recursion when data collected for an algorithm leads to real-world responses that feedback to the algorithm. For example, simulations of PredPol software (PredPol) used in Oakland, California, suggested increased police presence in black neighborhoods based on crime data reported by the public. The simulation showed that the public reported crimes based on the sight of police cars, regardless of what the police were doing. The simulation interpreted the sightings of police cars in modeling their crime predictions and, in turn, assigned an even more significant police presence in those neighborhoods. The Human Rights Data Analysis Group, which ran the simulation, cautioned that such feedback loops could reinforce and perpetuate racial discrimination in policing in places where racial discrimination is a factor in arrests. Another well-known example of an algorithm that engages in such behavior is COMPAS, software that determines the likelihood that a person will become a felon. The software is often criticized for being much more likely to classify blacks as criminals than others and then feeding the data back into itself when a person becomes a criminal, reinforcing the bias created by the data set to which the algorithm responds.

## ***Impacts of Algorithmic Bias***

### **A few examples:**

#### **Gender Discrimination**

In 2016, it was noted that the professional network LinkedIn recommends male variants of female names in search queries. However, the website did not give similar recommendations when searching for male characters. For example, searches for "Andrea" asked if users meant "Andrew," but searches for "Andrew" did not ask if users wanted to find "Andrea." The company said this resulted from an analysis of users' interactions with the site.<sup>vii</sup>

In 2012, the department store company Target was sued for collecting data points that could be used to infer when customers were pregnant, even if they had not announced it, and then sharing that information with marketing partners. Because the data was predicted and not directly observed or reported, the company was under no legal obligation to protect the privacy of these customers.<sup>viii</sup>

Web search algorithms are also accused of bias. For example, Google's results can favor pornographic content for search terms related to sexuality, such as "lesbian." This bias goes so far that the search engine displays popular but sexualized content for neutral search queries.

For example, "Top 25 Sexiest Women Athletes" articles are displayed on the first page when searching for "women athletes."<sup>ix</sup>

In 2017, Google adjusted these results and others that showed hate groups, racist views, child abuse, pornography, and other disturbing and offensive content.<sup>x</sup> Other examples include displaying better-paying jobs for male applicants on job search websites.<sup>xi</sup>

### Discrimination based on race and ethnic origin

Algorithms have been criticized as a method of masking racial bias in decision-making. However, because of the way certain racial and ethnic groups have been treated in the past, data can often contain hidden biases. For example, blacks are likely to receive longer sentences than whites for the same offense.<sup>xii</sup> This could mean that a system is reinforcing the original prejudices in the data.

In 2015, Google apologized when black users complained that an image recognition algorithm in the Photos application identified them as gorillas.<sup>xiii</sup> In 2010, Nikon cameras were criticized because image recognition algorithms asked Asian users if they blinked. Such examples are the result of biases in biometric datasets. Biometric data is derived from aspects of the body,

including observed or inferred racial characteristics, which can then be translated into data points. For example, speech recognition technology can have varying accuracies depending on the user's accent. This may be due to a lack of training data for speakers of that accent.<sup>xiv</sup>

Biometric data on race can also be inferred rather than observed. For example, a 2012 study showed that names commonly associated with blacks were more likely to lead to search results indicating arrests, regardless of whether police recorded the person's name.<sup>xv</sup> A 2015 study also found that blacks and Asians are assumed to have worse lung function because racial and occupational exposure data are not included in the lung function prediction algorithm model.<sup>xvi</sup>

In 2019, a research study found that a healthcare algorithm sold by Optum favors white patients over sick black patients. The algorithm predicts how much patients would cost the health care system in the future. However, the costs are not race-neutral, as black patients incurred about \$1,800 less in medical costs per year than white patients with the same number of chronic conditions, resulting in the algorithm rating white patients at the same risk for future health problems as black patients who suffered from significantly more diseases.<sup>xvii</sup>



A study conducted by UC Berkeley researchers in November 2019 found that mortgage algorithms discriminated against Latino and African Americans, which discriminated against minorities based on "creditworthiness," which is enshrined in the U.S. Fair Lending Act that allows lenders to determine whether a person is creditworthy based on identifying measures. These particular algorithms were present in FinTech companies and were shown to discriminate against minorities.<sup>xviii</sup>

### Commercial influences

Corporate algorithms could be biased to invisibly favor financial agreements or collusion between companies without the user's knowledge, who might believe the algorithm to be impartial. For example, American Airlines developed a flight search algorithm in the 1980s. The software presented customers with various flights from different airlines but weighed factors that favored its flights, regardless of price or convenience. Before the U.S. Congress, the airline's president said the system was developed to gain a competitive advantage through preferential treatment.<sup>xix</sup>

In a 1998 paper describing Google, the company's founders had adopted a policy of transparency in search

results concerning paid placement, arguing that "ad-supported search engines will be inherently biased toward advertisers and away from consumer needs." This bias, they claim, is an "invisible" manipulation of the user.<sup>xx</sup>

### Voting behavior

A series of studies of undecided voters in the U.S. and India found that search engine results can influence election outcomes by about 20%. The researchers concluded that candidates "have no way to compete" when an algorithm - with or without intent - raises page listings for a competing candidate. In addition, Facebook (meta) users who saw news related to the election were more likely to vote.<sup>xxi</sup> A 2010 randomized study of Facebook users found a 20% increase in turnout (340,000 votes) among users who saw messages encouraging voting and pictures of their friends who had voted.<sup>xxii</sup> Legal scholar Jonathan Zittrain warned that this could lead to a "digital gerrymandering" effect in elections, i.e., selective presentation of information by an intermediary pursuing its agenda rather than serving its users when intentionally manipulated.<sup>xxiii</sup>

### Law enforcement and litigation

Algorithms already have numerous applications in legal systems. One example is COMPAS, a commercial program

widely used by U.S. courts to assess a defendant's likelihood of recidivism. ProPublica claims that the average recidivism risk of black defendants as determined by COMPAS is significantly higher than the moderate risk of white defendants as determined by COMPAS. Black defendants are twice as likely to be incorrectly classified as "high risk" as white defendants.<sup>xxiv</sup>

A study, "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact," asserts that black defendants are twice as likely as white defendants to be classified as higher risk (45 percent versus 23 percent), even though they objectively did not recidivate over a two-year observation period.<sup>xxv</sup>

### Online hate speech

In 2017, a Facebook (Meta) algorithm designed to remove hate speech on the Internet was found to favor white males over black children when rating offensive content, according to internal Facebook documents.<sup>xxvi</sup> The algorithm, a combination of computer programs and human content reviewers, was designed to protect broad categories, not just specific subsets of types. So, for example, posts denouncing "Muslims" would be blocked, while posts criticizing "radical Muslims" would be allowed. An unexpected consequence of the algorithm is that hate

speech against black children is permitted because it denounces the "children" subgroup of blacks rather than "all blacks," while "all white males" would trigger blocking because whites and males are not considered subgroups.<sup>xxvii</sup> Facebook (Meta) also allowed ad buyers to target "Jew-haters" as a user category, which the company said was an unintended result of algorithms used to score and categorize data. The company's design also allowed ad buyers to exclude African Americans from viewing housing ads.<sup>xxviii</sup>

While algorithms are used to detect and block hate speech, some algorithms were found to flag information posted by black users as hate speech at 1.5 times the likelihood and flag information written in ebonics as such at 2.2 times the likelihood.<sup>xxix</sup> In addition, slurs and epithets were flagged without context, even when used by communities that reappropriated them.<sup>xxx</sup>

## Surveillance

Surveillance camera software can be seen as inherently political, requiring algorithms to distinguish normal from abnormal behavior and determine who belongs in certain places.<sup>xxxi</sup> The ability of such algorithms to recognize faces within a racial spectrum is limited by the racial diversity of the images in the training database; if the majority of the

photos belong to one race or gender, the software is better able to recognize other members of that race or gender.<sup>xxxii</sup> However, even audits of these image recognition systems are ethically questionable. Some scholars have pointed out that the context of the technology will always have a disproportionate impact on communities whose actions are overly monitored. A 2002 analysis of software used to identify people in CCTV images found several examples of bias in matching against crime databases. The software identified men more often than women, older people more often than young people, Asians, African Americans, and other races more often than whites.<sup>xxxiii</sup> Further studies of facial recognition software have found that the opposite is true when the software is trained on non-criminal databases, with the software being the least accurate in identifying dark-skinned women.<sup>xxxiv</sup>

## ***Conclusion***

### Conclusion

Algorithmic bias is a critical issue that requires immediate attention and action. Algorithms, while powerful tools are not free from biases inherent in the data on which they are

trained and in the decisions of their developers. The consequences of algorithmic bias are far-reaching, affecting individuals, communities, and society.

It is essential to be clear that algorithmic bias is not a problem that will solve itself. Left untreated, biases can perpetuate harmful stereotypes, exacerbate social inequalities, and perpetuate discrimination in various areas, including hiring, lending, criminal justice, and health care.

To defuse algorithmic bias, we need a multifaceted approach.

- First, we must prioritize diverse and inclusive data collection. We can reduce the risk of reinforcing existing biases by ensuring that the datasets used to train algorithms are comprehensive, balanced, and reflective of the population.
- Second, transparency and accountability are critical. Developers and organizations responsible for creating and deploying algorithms must be transparent about their methods and actively assess the potential for bias during development. Rigorous testing and ongoing reviews can help

identify and eliminate biases before they cause harm.

- Third, diversity in AI and technology is critical. By fostering diverse teams and perspectives, we can combat unconscious bias and develop more inclusive algorithms that cover a broader range of experiences and needs.

Regulatory action is also critical to ensure the fairness of algorithms. Government agencies and policymakers must work with experts, stakeholders, and the public to establish policies and regulations that promote equity and prevent discrimination in algorithmic systems.

In addition, we must embrace the concept of "explainable AI." Algorithms should not be treated as black boxes but should provide transparent explanations for their decisions. In this way, individuals can understand and challenge the outcomes of AI systems, especially if they feel they have been mistreated or excluded.

**Addressing algorithmic bias requires collaboration among technology developers, policymakers, researchers, and society. Only through continued education, collaboration, and vigilance can we build a future in which algorithms serve as equitable tools that**

support our efforts while upholding the values of fairness, justice, and equity.

By recognizing the importance of algorithmic bias and working together to address it, we can harness AI's transformative potential for humanity's benefit and create a world where technology empowers us all to thrive without discrimination or prejudice.

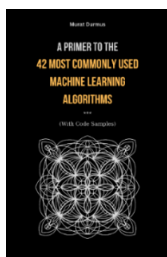
The path to fair and unbiased algorithms is challenging but worth pursuing because it promises a more just and equitable future for all. Let's move forward together, guided by empathy, ethics, and a commitment to harnessing the power of AI for the common good.



**MORE BOOKS BY THE AUTHOR**

# A PRIMER TO THE 42 MOST COMMONLY USED MACHINE LEARNING ALGORITHMS BOOK

(WITH CODE SAMPLES)



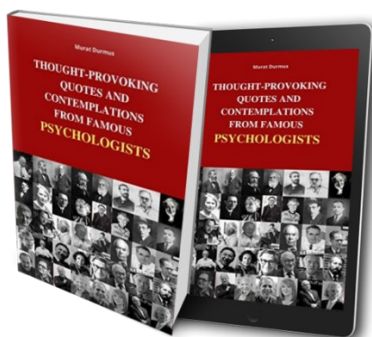
Whether you're a data scientist, software engineer, or simply interested in learning about machine learning, "A Primer to the 42 Most commonly used Machine Learning Algorithms (With Code Samples)" is an excellent resource for gaining a comprehensive understanding of this exciting field.

**Available on Amazon:**

<https://www.amazon.com/dp/B0BT911HDM>

THOUGHT-PROVOKING QUOTES & CONTEMPLATIONS FROM FAMOUS  
PSYCHOLOGISTS

# THOUGHT-PROVOKING QUOTES & CONTEMPLATIONS FROM FAMOUS PSYCHOLOGISTS



**(Over 600 Quotes & Contemplations)**

**Available on Amazon:**

<https://www.amazon.com/dp/B09NQ2BM1Y>

Kindle: **(ASIN: B09B79KR7P)**

Paperback: **(ISBN-13: 979-8543952337)**

# MINDFUL AI

## *Reflections on Artificial Intelligence*

Inspirational Thoughts & Quotes on Artificial Intelligence

(Including 13 illustrations, articles & essays for the  
fundamental understanding of AI)

***The field of AI is highly interdisciplinary & evolutionary. The  
more AI penetrates our life and environment, the more  
comprehensive the points we have to consider and adapt.  
Technological developments are far ahead of ethical &  
philosophical interpretations; this fact is disturbing.***

***We need to close this gap as soon as possible.***

~ (Mindful AI)



Available on Amazon:

<https://www.amazon.com/dp/B0BKMK6HLJ>

# REFERENCES

- 
- <sup>i</sup> Confirmation bias. (2023, June 25). In Wikipedia. [https://en.wikipedia.org/wiki/Confirmation\\_bias](https://en.wikipedia.org/wiki/Confirmation_bias)
- <sup>ii</sup> Sunk cost. (2023, June 14). In Wikipedia. [https://en.wikipedia.org/wiki/Sunk\\_cost#Fallacy\\_effect](https://en.wikipedia.org/wiki/Sunk_cost#Fallacy_effect)
- <sup>iii</sup> Gambler's fallacy. (2023, June 30). In Wikipedia. [https://en.wikipedia.org/wiki/Gambler%27s\\_fallacy](https://en.wikipedia.org/wiki/Gambler%27s_fallacy)
- <sup>iv</sup> Dunning–Kruger effect. (2023, June 23). In Wikipedia. [https://en.wikipedia.org/wiki/Dunning%E2%80%93Kruger\\_effect](https://en.wikipedia.org/wiki/Dunning%E2%80%93Kruger_effect)
- <sup>v</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- <sup>vi</sup> Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- <sup>vii</sup> Day, Matt (31 August 2016). "How LinkedIn's search engine may reflect a gender bias". *The Seattle Times*. Retrieved 25 November 2017.
- <sup>viii</sup> Crawford, Kate; Schultz, Jason (2014). "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms". *Boston College Law Review*. **55** (1): 93–128. Retrieved 18 November 2017.
- <sup>ix</sup> Noble, Safiya (2012). "Missed Connections: What Search Engines Say about Women" (PDF). *Bitch*. **12** (4): 37–41.
- <sup>x</sup> Guynn, Jessica (16 March 2017). "Google starts flagging offensive content in search results". *USA TODAY*. USA Today. Retrieved 19 November 2017.

- 
- <sup>xi</sup> Simonite, Tom. "[Study Suggests Google's Ad-Targeting System May Discriminate](#)". *MIT Technology Review*. Massachusetts Institute of Technology. Retrieved 17 November 2017.
- <sup>xii</sup> Alexander, Rudolph; Gyamerah, Jacquelyn (September 1997). "Differential Punishing of African Americans and Whites Who Possess Drugs: A Just Policy or a Continuation of the Past?". *Journal of Black Studies*. **28** (1): 97–111. doi:[10.1177/002193479702800106](#). ISSN [0021-9347](#). S2CID [152043501](#).
- <sup>xiii</sup> Gynn, Jessica (1 July 2015). "[Google Photos labeled black people 'gorillas'](#)". *USA TODAY*. USA Today. USA Today. Retrieved 18 November 2017.
- <sup>xiv</sup> "Alexa does not understand your accent". *Washington Post*.
- <sup>xv</sup> Sweeney, Latanya (28 January 2013). "Discrimination in Online Ad Delivery". [arXiv:1301.6822 \[cs.LR\]](#).
- <sup>xvi</sup> Braun, Lundy (2015). "[Race, ethnicity and lung function: A brief history](#)". *Canadian Journal of Respiratory Therapy*. **51** (4): 99–101. ISSN [1205-9838](#). PMC [4631137](#). PMID [26566381](#).
- <sup>xvii</sup> Johnson, Carolyn Y. (24 October 2019). "[Racial bias in a medical algorithm favors white patients over sicker black patients](#)". *Washington Post*. Retrieved 2019-10-28.
- <sup>xviii</sup> Bartlett, Robert; Morse, Adair; Stanton, Richard; Wallace, Nancy (June 2019). "[Consumer-Lending Discrimination in the FinTech Era](#)". *NBER Working Paper No. 25943*. doi:[10.3386/w25943](#). S2CID [242410791](#).
- <sup>xix</sup> Sandvig, Christian; Hamilton, Kevin; [Karahalios, Karrie](#); Langbort, Cedric (22 May 2014). "[Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms](#)" (PDF). *64th Annual Meeting of the International Communication Association*. Retrieved 18 November 2017.
- <sup>xx</sup> Brin, Sergey; Page, Lawrence. "[The Anatomy of a Search Engine](#)". [www7.scu.edu.au](#). Archived from [the original](#) on 2 July 2019. Retrieved 18 November 2017.
- <sup>xxi</sup> Epstein, Robert; Robertson, Ronald E. (18 August 2015). "[The search engine manipulation effect \(SEME\) and its possible impact on the outcomes of elections](#)". *Proceedings of the National Academy of*

- Sciences*. **112** (33): E4512–E4521. [Bibcode:2015PNAS..112E4512E](#). [doi:10.1073/pnas.1419828112](#). [PMC 4547273](#). [PMID 26243876](#).
- <sup>xxii</sup> Bond, Robert M.; Fariss, Christopher J.; Jones, Jason J.; Kramer, Adam D. I.; Marlow, Cameron; Settle, Jaime E.; Fowler, James H. (13 September 2012). ["A 61-million-person experiment in social influence and political mobilization"](#). *Nature*. **489** (7415): 295–8. [Bibcode:2012Natur.489..295B](#). [doi:10.1038/nature11421](#). [ISSN 0028-0836](#). [PMC 3834737](#). [PMID 22972300](#).
- <sup>xxiii</sup> Zittrain, Jonathan (2014). ["Engineering an Election"](#) (PDF). *Harvard Law Review Forum*. **127**: 335–341. Retrieved 19 November 2017.
- <sup>xxiv</sup> Jeff Larson, Julia Angwin (2016-05-23). ["How We Analyzed the COMPAS Recidivism Algorithm"](#). *ProPublica*. [Archived](#) from the original on 29 April 2019. Retrieved 2020-06-19.
- <sup>xxv</sup> Skeem J, Lowenkamp C, Risk, Race, & Recidivism: Predictive Bias and Disparate Impact, (June 14, 2016). Available at SSRN: <https://ssrn.com/abstract=2687339> or <https://doi.org/10.2139/ssrn.2687339>
- <sup>xxvi</sup> Angwin, Julia; Grassegger, Hannes (28 June 2017). ["Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children — ProPublica"](#). *ProPublica*. Retrieved 20 November 2017.
- <sup>xxvii</sup> Angwin, Julia; Grassegger, Hannes (28 June 2017). ["Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children — ProPublica"](#). *ProPublica*. Retrieved 20 November 2017.
- <sup>xxviii</sup> Angwin, Julia; Varner, Madeleine; Tobin, Ariana (14 September 2017). ["Facebook Enabled Advertisers to Reach 'Jew Haters' — ProPublica"](#). *ProPublica*. Retrieved 20 November 2017.
- <sup>xxix</sup> Sap, Maarten. ["The Risk of Racial Bias in Hate Speech Detection"](#) (PDF). [Archived](#) (PDF) from the original on 2019-08-14.

- 
- <sup>xxx</sup> Ghaffary, Shirin. "[The algorithms that detect hate speech online are biased against black people](#)". Vox. Retrieved 19 February 2020.
- <sup>xxxi</sup> Graham, Stephen D.N. (July 2016). "[Software-sorted geographies](#)" (PDF). *Progress in Human Geography* (Submitted manuscript). **29** (5): 562–580. [doi:10.1191/0309132505ph568oa](#). [S2CID 19119278](#).
- <sup>xxxii</sup> Furl, N (December 2002). "[Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis](#)". *Cognitive Science*. **26** (6): 797–815. [doi:10.1207/s15516709cog2606\\_4](#).
- <sup>xxxiii</sup> Introna, Lucas; Wood, David (2004). "[Picturing algorithmic surveillance: the politics of facial recognition systems](#)". *Surveillance & Society*. **2**: 177–198. Retrieved 19 November 2017.
- <sup>xxxiv</sup> Buolamwini, Joy; Gebru, Timnit (2018). "[Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification](#)" (PDF). *Proceedings of Machine Learning Research*. **81**: 1 – via MLR Press.