

# Supplemental Material for *Falsehoods Fly*

## Chapter 1: Lies Kill: The Perils of Misinformation

### Primitive Information

The AIMS theory of information concerns the rich, meaningful kinds of information that operate in humans, some animals, and the most sophisticated computers. Such semantic information can turn out to be false (misinformation), and can also turn out to be lies (disinformation). My book *Bots and Beasts* provides a systematic comparison of human intelligence with that of nonhuman animals and current computers.<sup>1</sup> Although human intelligence surpasses that of animals and computers, the AIMS theory could easily be applied to the smartest animals, such as birds capable of deception, and to the most advanced current examples of artificial intelligence, such as IBM Watson.

But much simpler kinds of information have also been ascribed to physical and biological systems.<sup>2</sup> Here are some examples of information that I call “primitive” because they do not have the rich structure, meaning, and purpose of human communication.

- Physical information operates in any system that can be described by probabilities, for example matter falling into a black hole.
- Quantum information concerns the state of systems at atomic or sub-atomic levels, for example when a logic gate in a quantum computer can be on, off, or both.
- Genetic information is stored in an organism’s DNA molecules and provides instructions for making proteins.
- Bacterial information consists of molecular signals sent between single-celled organisms in response to nutrients and hazards.

- Plant information consists of chemical signals sent between organisms, for example when trees communicate via roots and fungi.

These uses of the word “information” are appropriate in their contexts, but should not be confused with the kind of information needed to explain human problems with misinformation and disinformation.

The AIMS theory of information serves that explanatory purpose and can ignore much simpler systems where acquisition, inference, memory, and spread are absent or deficient in syntax, semantics, and pragmatics. Syntactic complexity requires relational structure that can discriminate, for example, between *dog chases person* and *person chases dog*. Such relations can be represented linguistically by grammar but also in other formats such as pictures and auditory images. Semantic complexity requires being able to connect the syntactic structures to each other and to the world, for example in knowing what it means for a real dog to chase a real person. Moreover, pragmatic complexity allows information to contribute to the accomplishment of the goals of an organism, which in humans include not just survival and reproduction but also flourishing in love, work, and play. The AIMS theory applies to humans and other entities capable of abundant representations, meanings, and purposes. Stars, atoms, DNA, bacteria, and plants need different theories, perhaps based on probabilities rather than on mechanisms for acquisition, inference, memory, and spread.<sup>3</sup>

## **Chapter 2: Information and Misinformation: How They Work**

### **Interactions of Mechanisms**

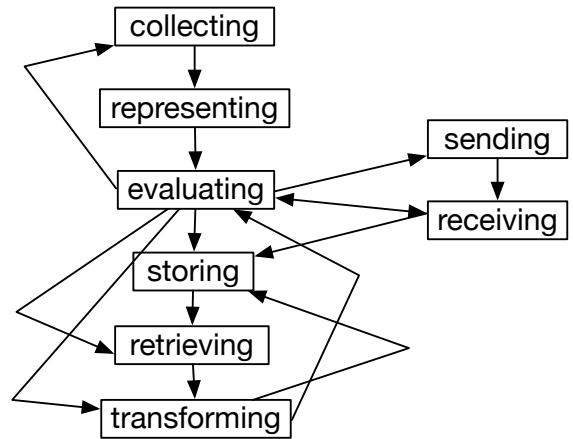
In biological systems, mechanisms often interact, for example when digestion provides energy for the respiratory system that furnishes oxygen to the cardiovascular

system. Moreover, diseases can result from interaction among broken mechanisms, for example when congestive heart failure leads to fluid in the lungs that produces shortness of breath, increasing strain on the heart. Similarly, well-functioning information systems gain from fruitful interactions of mechanisms concerning acquisition, inference, memory, and spread. Moreover, breakdowns of these mechanisms can interact to produce cascades of misinformation. Interactions of mechanisms are also important for understanding techniques of reinformation and interconnections among domains of misinformation such as pandemics and climate.

### *Interactions among Information Mechanisms*

The development and spread of information have been triumphs of science and technology, leading quickly to novel explanations, treatments, and preventions. These advances result not just from the sum of the 8 mechanisms for information but from their productive interactions.

Figure S.1 illustrates how collecting leads to representing which leads to evaluating, which feeds back into collecting and forward to storing, transforming, sending and receiving. Effective development and use of information requires these 8 mechanisms to interact smoothly. For example, Pat collects information about Sam by perception, represents it as words or images, and evaluates it for importance and credibility. The centrality of evaluating is clear from the role that it is supposed to play in triggering cognitive processes of memory and inference, as well as social processes of communication.

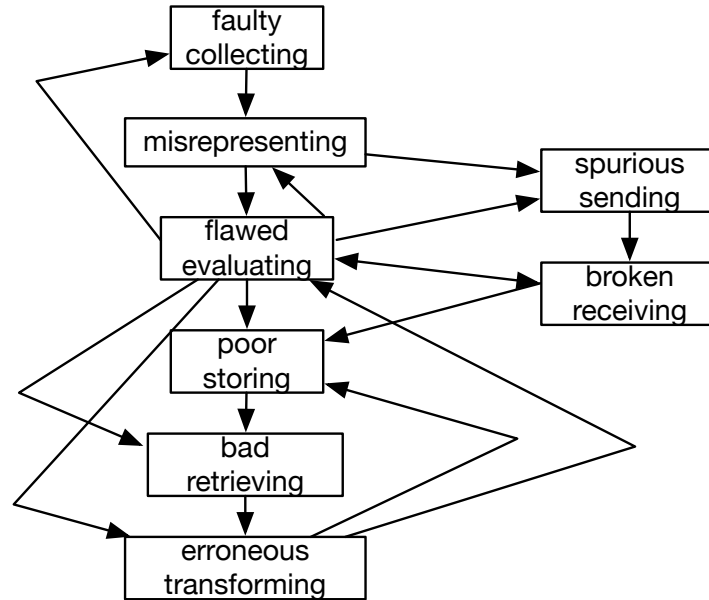


**Figure S.1** Interactions among mechanisms of information.

*Interactions of Mechanism Breakdowns*

Breakdowns in these interactions among mechanisms can also be sources of misinformation, in addition to breakdowns in the particular mechanisms. In medicine, comorbidity is the presence of more than one condition that diminishes the health of a patient. Different mechanism breakdowns can interact to cause even more harm, for example when a COVID-19 patient suffers from both respiratory problems and blood clots. Similarly, misinformation can result from interactions among multiple breakdowns, as illustrated in figure S.2 which adapts the information mechanisms of figure S.1.

For example, Pat may make a perceptual error in thinking that Sam has green eyes when they are actually brown. This faulty collecting can lead to misrepresenting in the verbal false belief that Sam’s eyes are green or with the mistaken visual image of green eyes. If Pat is overconfident in this judgment and fails to evaluate it, then Pat will store the false information and retrieve it later to lead to erroneous inferences. Moreover, Pat may tell others about Sam’s green eyes leading to the spread of misinformation. Hence misinformation comes not just from breakdowns in the individual mechanisms that produce information, but also from deviant interactions among the mechanisms.



**Figure S.2** Interactions among mechanism breakdowns that foster misinformation.

Figure S.2 can sadly be read as a recipe for how to spread misinformation by exploiting breakdowns in information mechanisms that could otherwise be powerful means of developing and communicating well-founded ideas. Fortunately, broken mechanisms can sometimes be repaired.

### **Interactions among Reinforcement Techniques**

An important question concerns the extent of interactions among the techniques of correcting misinformation listed in chapter 1. Spotting misinformation, locating sources, and identifying their motives are preliminaries for the rectification methods of factual correction, critical thinking, motivational interviewing, institutional modification, and political action.

Sometimes these methods can work together: political action depends on getting voters to choose informed leaders, which requires educating the individual voters and leaders using factual correction, critical thinking, and motivational interviewing. Factual

correction can contribute to critical thinking by making people aware that they have been misinformed and should look for further errors that can be diagnosed and corrected. Motivational interviewing might serve as an initial, gentle way of making people aware of domains that require critical thinking. For example, anti-vaxxers approached using empathic questioning may come to realize that they have been misled by bogus authorities who deserve critical examination. Institutional modification in groups such as social media companies and political parties could facilitate political action.

### *Interactions among Domain Areas*

A final group of important interactions concerns the domains in which misinformation has been rampant, including science, medicine, politics, social inequality, and war. Chapters 4-8 mostly consider these domains in isolation from each other, but causal connections can establish relationships among their episodes of misinformation. For example, increasingly serious climate change leads to deforestation and other environmental degradation, which leads to the encroachment of humans on animals such as bats that harbor viruses that can spread diseases to humans and generate new pandemics. Political conspiracy theories such as claims of Jewish plans to dominate the world also serve as myths to justify social inequality. The COVID-19 pandemic hindered public education and thereby increased inequalities in learning by children with poorer access to technology.

Such interactions among domains of misinformation imply that reinformation cannot be piecemeal but must often deal with large packages of mistakes. At the extreme, some people are systematically anti-vaxxers, science skeptics, conspiracy theorists, racists,

*and* warmongers. Although based on misinformation, these views have a high degree of cognitive and emotional coherence that can make them extraordinarily hard to dislodge.

### **Why These Mechanisms?**

I am proposing to explain information and misinformation by the four AIMS processes that are fleshed out by the eight classes of mechanisms. What justifies these particular mechanisms as central to a theory of information? Acceptance of the AIMS mechanisms is based on their psychological and social reality, superiority to alternative accounts of information, applicability to real-world misinformation domains, and ability to explain empirical findings about misinformation and reinformation.

First, evidence is ample that each of these mechanisms operates in human minds and social interactions. The role of perception in collecting information has been documented by countless studies in psychology and neuroscience. The history of science thoroughly illustrates the role of instruments, systematic observations, and experiments in the collection of information. Representing has been a central part of cognitive theorizing since the demise of behaviorism in the 1950s and 1960s, with many studies concerning the operation of verbal and imagistic representations.<sup>4</sup> Similarly, inferential mechanisms of evaluating and transforming have been thoroughly investigated in cognitive psychology, as have memory mechanisms of storing and retrieving. Finally, the sending and receiving mechanisms of information spread are familiar from many studies of human and electronic communication. My description of the 8 mechanisms may seem qualitative and vague, but all of them have been specified with sufficient rigor to produce computational models which generate simulations to approximate human behavior.<sup>5</sup>

Second, the AIMS theory of information and misinformation provides more detailed and mechanistic explanations than previous accounts based on probability and philosophical analysis. Claude Shannon (1948) developed a mathematical theory of communication that has become the basis for what is now widely known as “information theory”.<sup>6</sup> He said that a communication system has an information source that transmits messages using signals sent to a receiver that reconstructs the message for the person or thing for whom the message is intended. Information is understood statistically on the assumption that the most informative messages are ones that occur with the lowest probability. The amount of information  $H$  produced by a source is captured by the following equation where  $K$  is a positive constant and  $p_i$  is the probability of message  $i$ :

$$H = -K \sum p_i \log p_i.$$

Using logarithms to the base 2, the amount of information in a message  $m$  can be measured in bits by:  $I(m) = -\log p_m$ .

Shannon acknowledged that his account ignored the meaning of messages and critics have remarked that his neglect of semantics makes his equations a theory of signaling rather than of information.<sup>7</sup> Understanding the difference between information and misinformation depends on recognizing the difference in meaning between true and false claims. In Shannon’s engineering context, probability could naturally be understood as frequency of events such as a receiver sending a dot rather than a dash. But probability of sentences in real-world domains such as medicine requires different interpretations of probability as subjective degrees of belief or logical relations. Accordingly, probability does not furnish a theory of misinformation that challenges my mechanistic account.



In books and articles, the philosopher Luciano Floridi has provided a comprehensive and penetrating discussion of the epistemology and ethics of information.<sup>8</sup> He describes a typical information life cycle which consists of creation, collection, recording, processing, distribution, consumption, and recycling, all processes which overlap with my 8 mechanisms of information. But he does not analyze these processes mechanistically in terms of their parts, connections, and interactions. Moreover, he does not describe misinformation as resulting from breakdowns in these processes, and he provides no general account of the origins and spread of misinformation.

Floridi defines information as consisting of data that are well-formed and meaningful, where a datum is something that is distinct from something else. He characterizes misinformation as false semantic content, in contrast to factual semantic information which is meaningful and *veridical*, a term that extends beyond true sentences to cover maps, diagrams, videos, and other kinds of representations. But he insists that false information is not information, in the same way that a false friend is not a friend, contrary to the terminology of Surgeon General Murthy and many others who take misinformation to be false information.

For semantic information, I prefer to construe pieces of information as representations rather than data to capture their function of standing for aspects of the world including non-observable claims about causality. Floridi's characterization of data in terms of distinctiveness is unclear, and further confusion arises from the way many dictionaries define data as pieces of information.

Cailin O'Connor and James Owen Weatherall provide a sustained and illuminating discussion of misinformation across various fields, but their emphasis on false beliefs

ignores non-sentential forms of misinformation.<sup>9</sup> They use computer models of belief spread as mechanisms for misinformation, and the models yield some interesting results about how easily beliefs lacking in evidence can spread through a community. But the models are drastically oversimplified, taking agents as using Bayesian rationality applied to sentences, with no considerations of the contribution to misinformation of vivid images, emotions, motivations, flawed evaluations, and erroneous inferences. Moreover, the models ignore the complexities of storing, retrieving, sending and receiving representations. Hence the scope of the account of misinformation by O'Connor and Weatherall is much narrower than my mechanistic theory.

In the earlier 1950s, psychologists explored the relevance of Shannon's mathematical theory of communication for understanding mental processes in terms of bits of information.<sup>10</sup> They soon realized the limitations of probabilistic understanding of information and shifted to considerations of chunks, representations, and processes analogous to computer programs.<sup>11</sup> My account of the mechanisms of information is compatible with this tradition but expands it to include details about social mechanisms of sending and receiving. My brief review of alternative theories of information suffices to show that the AIMS theory has major advantages over previous accounts, especially for explaining misinformation.

The third and major reason for recognizing the eight mechanisms in the AIMS theory as successful explanations of information and misinformation is their application to five important domains: medicine, science, politics, society, and international relations. Each of these domains receives detailed treatment in chapters 4-8 which show the rich applicability of the AIMS theory to real-world problems.

Finally, the fourth reason for recognizing the eight mechanisms in the AIMS theory is that they serve to explain recent empirical findings about misinformation. I now review several such findings and connect them with AIMS mechanisms.

### **Empirical Findings About Misinformation**

Psychologists have turned their attention to “misinformation processing” through studies of how individuals operate with false beliefs.<sup>12</sup> In psychology, an effect is a statistically significant relationship between observed factors, for example practice findings that spending time on a task leads to improved future performance. Psychological research has uncovered effects for both misinformation and reinformation.

#### *Misinformation Effects*

Various psychological phenomena describe how people are inclined to accept misinformation. All of these can be explained as breakdowns in AIMS mechanisms.

The *illusory truth effect* is that people tend to judge repeated information to be true just because it is familiar.<sup>13</sup> It could also be called the *repetition effect*, and is important for understanding the unfortunate efficacy of ignoring reality and instead getting people to accept made-up claims just by frequently reasserting them. For example, Pat may believe a rumor reported by both Sam and Quinn without realizing that Quinn heard it from Sam. The primary mechanism breakdown behind this effect is bad receiving that exploits people’s default acceptance of what they are told. Also implicated is flawed evaluating because people fail to notice that the information emanates from spurious sending and that apparent corroboration from multiple sources is as bogus as being convinced by reading multiple copies of the same newspaper edition.

The illusory truth effect is a special case of the mere-exposure effect, in which people tend to develop a preference for things merely because they are familiar with them.<sup>14</sup> The mere-exposure effect also results from flawed evaluating due to brain interconnections between cognition and emotion described in chapter 3.

The *motivated reasoning effect* is that people are inclined to believe what fits with their personal goals and to distort evidence in the service of those goals.<sup>15</sup> Such inference operates in everyone's thinking as we all tend to believe what we want to believe concerning personally important concerns such as health, success, and personal relationships. Motivated reasoning results from flawed evaluating that can occur because of biased memory and emotions, as chapter 3 details.

An important special case of motivated reasoning is the *identity effect*, which is the tendency of people to make decisions and form beliefs based on who they think they are and what they want to become.<sup>16</sup> Rather than think through large amounts of information, people can more easily interpret situations based on their own social identities. For example, Pat and Sam may strongly identify as Canadians and therefore believe that ice hockey is the world's best sport.

The *confirmation bias effect* is the tendency of people to find, recall, and prefer information that supports their prior beliefs.<sup>17</sup> Sometimes confirmation bias is an instance of motivated reasoning when people's goals include preserving their prejudices, but confirmation bias can operate more generally through unjustified avoidance of belief change. Preferring maintenance of existing beliefs over new, better supported ones is another case of flawed evaluating.

In the book, I discuss another misinformation effect, confusedly called *the* misinformation effect, which occurs when memory recall becomes ineffective because of confusion with later information. This effect results from the mechanism breakdowns of poor storing and bad retrieving.

These psychological effects can interact, for example when bullshitters combine the illusory truth effect and the motivated reasoning effect by repeating what people want to hear to support their personal goals. Critical thinking should recognize that some beliefs are supported by more than one error tendency.

### *Reinformation Effects*

Reinformation is the process of overcoming misinformation by restoring valid information. Psychological research has identified why reinformation is hard but also why it can sometimes be effective.

The *continued influence effect* is that people often rely on misinformation even after it has been corrected, a pernicious form of confirmation bias.<sup>18</sup> For example, Pat's memory may continue to contain the false information that Sam is a lawyer even after Quinn corrects the earlier message. This effect highlights the difficulty of accomplishing reinformation: even if people are told that they have acquired misinformation, they may stick with it even though they have reason to reject it. The main mechanism breakdowns responsible for the continued influence effect concern inference and memory. If people practiced ongoing evaluation of their beliefs, then those that do not fit with new information would be tossed out. But continuous evaluating would require more time and effort than most people want to apply, so inference does not succeed in throwing out misinformation.

Memory is also at fault in this effect, because we have no general way of removing stored beliefs from memory even if they have been retracted.

The *debunking effect* is that correcting misinformation can be effective if it uses detailed refutations that provide clear explanations of why the challenged information is false and what is true instead.<sup>19</sup> This effect is reassuring because it suggests that reinformation can sometimes overcome the continued influence effect. The underlying mechanism for the debunking effect is evaluating, carried out thoroughly so that people can grasp the need to abandon false beliefs.

Inoculation or prebunking is warning people not to acquire dubious information, for example if Pat tells Sam not to believe anything that Quinn says. The *inoculation effect* is that prior warning is more effective than debunking: it is easier to warn someone away from misinformation than to correct it once incorporated.<sup>20</sup> The mechanistic explanation of debunking failure is similar to that of the continued influence effect, involving failures of evaluating and memory. Inoculation is helpful because it prompts people to do a better job of evaluating a claim before storing it in memory. Like cancer, misinformation is often easier to prevent than to cure, as chapter 8 discusses concerning preinformation.

Another encouraging empirical result about reinformation might be called the *accuracy nudge effect*. Researchers have found that subtly encouraging people to consider the accuracy of information that they transmit on social media can reduce the spread of misinformation.<sup>21</sup> The relevant broken mechanisms here are spurious sending and flawed evaluating: when nudged to be accurate, people become more careful about sending misinformation. As shown in figures 2.4 and 2.5, evaluating, sending, and receiving interact. One attempt to replicate the accuracy nudge effect with larger sample size did not

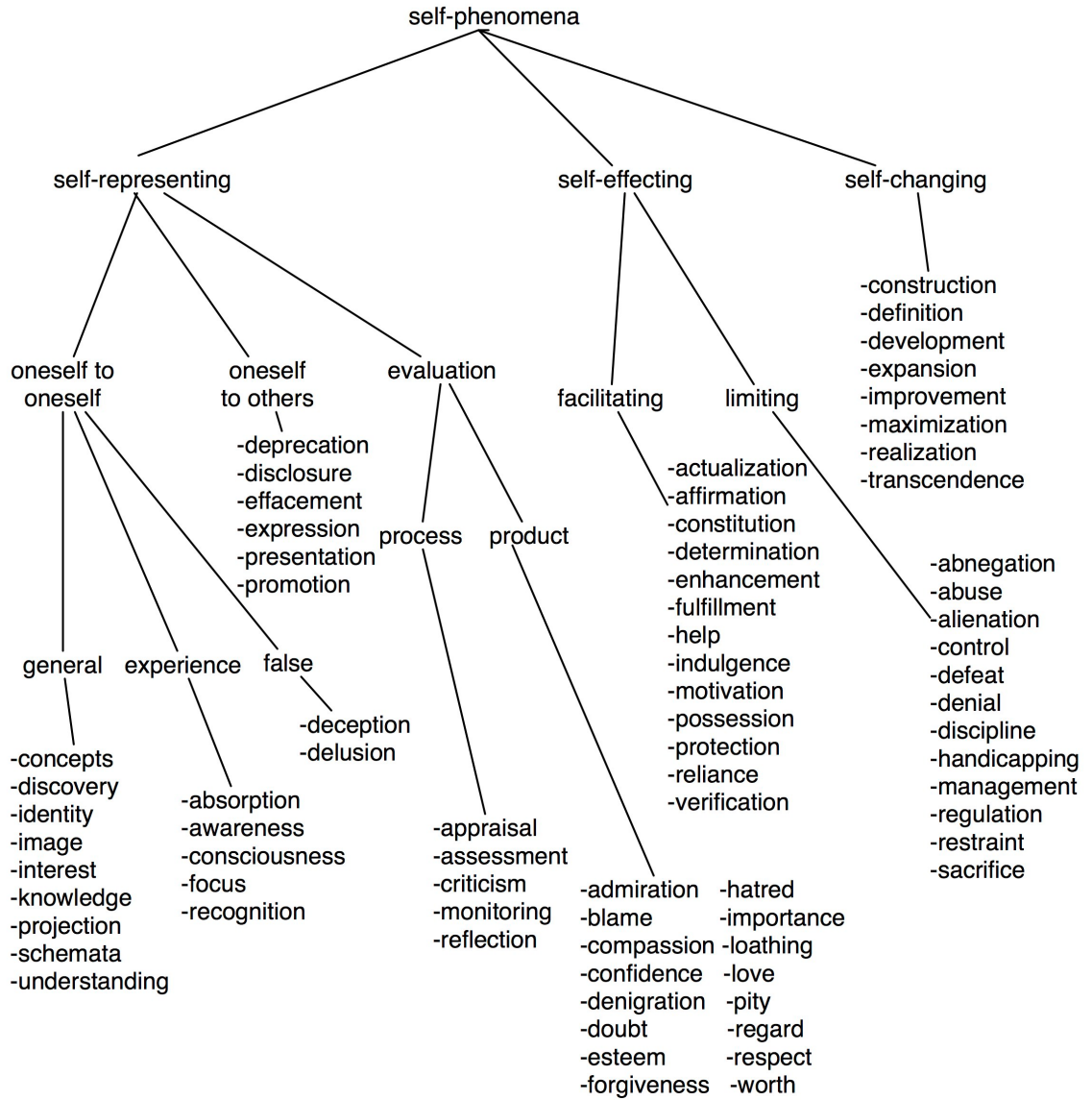
confirm it, so further investigation is needed to determine whether accuracy nudging is a good strategy for misinformation.<sup>22</sup>

A more discouraging potential effect that failed to replicate is the *backfire effect*, according to which attempts at reinformation only made matters worse by strengthening misconceptions. Fortunately, subsequent research found that corrective information can be effective although misinformation sometimes persists as suggested by the continued influence effect.<sup>23</sup>

Misinformation effects explain the tendencies of people to acquire inaccurate information, and reinformation effects explain the difficulties of disabusing them of their mistakes. I have reviewed these effects for two reasons. First, they illuminate aspects of<sup>24</sup>the course of misinformation in the five domains discussed in chapters 4-8. For example, the analysis of COVID-19 information in chapter 4 will apply motivated reasoning. Second, the misinformation and reinformation effects provide evidence for the mechanisms of the AIMS theory that explain the effects much more thoroughly than alternative theories based on probability or philosophical analysis. The major argument for the AIMS theory, however, depends on the theory's practical utility in the domains of medicine, science, politics, society, and international relations.

### **Chapter 3: Believing What You Want: Motivated Reasoning, Emotion, and Identity**

Identity requires an appreciation of the self, which involves the wide range of phenomena shown in figure S.3.



**Figure S.3** Grouping of self-phenomena. Source: Thagard 2019a.

## Chapter 4: Plagues: COVID-19 and Medical Misinformation

### Bayesian Alternative to Explanatory Coherence

The major alternative to the explanatory coherence account of causal inference uses probabilities calculated by Bayes' theorem.<sup>25</sup> But applying the Bayesian approach to real-life cases such as COVID-19 has numerous problems.<sup>26</sup> The required probabilities are



rarely known, such as the precise conditional probability of the evidence that states with mask requirements have less spread given the hypothesis that masks prevent COVID-19. More technically, the interpretation of such probabilities is unclear: are they frequencies or degrees of belief? The common Bayesian assumption that probabilities are degrees of belief is incompatible with experimental findings that human thinking often conflicts with the probability theory. Applying Bayes' inference to large sets of beliefs is computationally difficult as the number of calculations grows exponentially with the number of beliefs.

### Chapter 5: Storms: Climate Change and Scientific Misinformation

Here is a more detailed version of figure 5.4 prepared by Manjana Milkoreit.

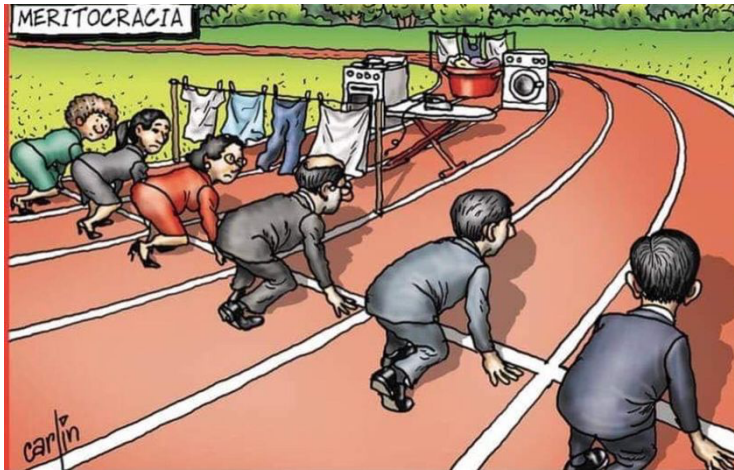


**Figure S.4** Map of values opposed to climate change. Ovals indicate desired values while hexagons indicate disliked values. Solid lines indicate

emotional associations while dotted lines indicate emotional incompatibility. The rectangle indicates neutrality, and the combined hexagon/oval indicates ambivalence. Source: Homer-Dixon et al. 2014, p. 15.

## Chapter 6: Plots: Conspiracies and Political Misinformation

## Chapter 7: Evils: Inequality and Social Misinformation



**Figure S.5** Inequality cartoon. Source: Carlin (Carlos Tovar Samanez). I could not use this in the book because I was unable to contact the cartoonist to request permission.

## Institutions and Systemic Racism

Based on: <https://www.psychologytoday.com/ca/blog/hot-thought/202301/what-systemic-racism-is-and-how-to-overcome-it>

Institutions are major contributors to real information and misinformation, which makes their modification a crucial target for reinformation. But what is an institution? It is more than just a collection of individuals, because institutions have properties such as policies and practices that go beyond the mental states of any person.<sup>27</sup> The assertion that “institutions are rules that govern people’s behavior in complex social situations” fails to

recognize that institutions can consist of other institutions as well as people, and to specify how rules govern behavior.<sup>28</sup>

A richer account comes from recognizing institutions as *systems* of interacting parts and properties that emerge from these interactions. Social problems are often described as “systemic”, for example in claims about systemic racism.<sup>29</sup> Nick Chater and George Loewenstein argue that social policies including ones dealing with misinformation have focused too much on individual behavior and neglected systemic reforms.<sup>30</sup> But what are the systems that are implicated and what is their relation to individuals?

We need an account of social systems that shows why they are relevant to explaining important phenomena such as racism and misinformation, and an account of individual and social explanations that sees them as complementary rather than competing. The term “institutional racism” is sometimes used as equivalent to “systemic racism”, so we need an account of the relation between systems and institutions.

The most rigorous characterization of systems I know is due to the philosopher Mario Bunge.<sup>31</sup> His CESM model of a system has it consisting of:

*Composition:* collection of all the parts of the system.

*Environment:* collection of items that act on components of the system.

*Structure:* collections of relations such as bonds among components or among these and items in its environment.

*Mechanism:* Collection of processes that makes the system behaves the way it does.

In social systems such as families and schools, the composition (set of parts) consists of people, but the parts can also be other social systems, for example when the United Nations includes numerous agencies such as the World Health Organization and UNESCO. The

environment consists of items in the world with which people interact such as the buildings in which they live or work. The structure of social system is all the relations among people such as being married and external relations such as living in a house. Bunge’s description of a mechanism as a collection of processes is consistent with but vaguer than my characterization of mechanisms as combinations of connected parts whose interactions produce regular changes.

Social systems operate by social mechanisms subject to the three-fold characterization in table S.1.

**Table S.1** 3-analysis of *social mechanism*.<sup>32</sup>

<i>Examples</i>	Families, clubs, political parties, markets, schools, international organizations, religions, hospitals, courts, schools, companies, clans, tribes, etc.
<i>Typical features</i>	People as parts, groups as wholes, groups as parts of larger groups, connections among people and groups, interactions through verbal and nonverbal communication, purpose or function, changes in individuals and groups.
<i>Explanations</i>	Explains: changes in people and groups. Explained by: human needs such as food and relationships.

All the examples in table S.1 are institutions, so perhaps all institutions can be understood as social mechanisms. But we cannot simply equate systems with mechanisms, because the usual account of mechanisms says nothing about environments. The behavior of mechanisms depends not only on the interactions of their parts but also on the causal effects on the parts from the environment in which the mechanism operates. For example,

brain mechanisms depend on sensory inputs to neurons that originate with external environments, for example when light reflected off a house stimulates retinal cells that send signals to the visual regions of the brain. The heart is a mechanism that pumps blood in part because the muscle cells in it get glucose through the body's digestion of food acquired through interaction with the world. So a better characterization of a system would be as a combination of environment, connected parts, wholes with emergent properties, interactions between parts, interactions between wholes, and changes that result from all these interactions.

How does this apply to systemic racism? Individual racism occurs when particular people have beliefs and emotions that are prejudicial against people of different races, for example when whites hate blacks or consider them inferior. Such localized racism might best be treated by changing the beliefs and emotions of the relevant individuals, through a combination of information, critical thinking, and motivational interviewing.

A social system is not just an aggregate of individuals because its operations depend on:

- (1) interactions between the individuals consisting of their communications with each other,
- (2) whole groups of individuals organized into institutions such as families, schools, police departments, hospitals, companies, and government agencies, and
- (3) environments in which the individuals operate including substandard housing, unavailability of good food stores, ineffective hospitals, etc.

Sometimes institutions are explicitly racist as indicated by their public policies and practices. Joe R. Feagin describes how the United States has had three stages of systemic

racism, two explicit and one implicit.<sup>33</sup> Slavery and legal segregation after the civil war both legally legitimized damaging treatment of Blacks with respect to employment, housing, education, housing, and transportation. Supreme Court decisions in the 1950s and 1960s eliminated such explicit discrimination.

But the operation of institutions is not determined purely by their explicit rule-based policies. Values, norms, routines, practices, and goals may not be part of the written policies of an institution for them to have causal influences on the behaviors of its members. Racism does not stop just because it is illegal, and the practices of institutions such as police departments, schools, companies, and housing agencies can continue despite legal changes.

How individuals communicate with each other depends on the institutions to which they belong, because groups such as families and workplaces provide channels. Institutions do not reduce to their members because they have emergent properties including explicit written-down policies and rules, and also implicit values and norms. Values are emotionally-charged concepts that emerge from the interactions of individuals. For example, to say that a country values democracy is more than that its individuals have a positive emotional attitude toward the concept of democracy. Individuals acquire these emotional attitudes in part through interaction with the agents of the country such as politicians and employees. Moreover, the value of democracy is instantiated in written policies such as constitutions and laws that decree free elections. Rules are written and mental structures of the form IF A THEN B, and norms are a special kind of rule that prescribes behaviors with an emotional commitment, as in: IF you lose an election, THEN admit defeat.<sup>34</sup>

To make matters even more complicated, social systems can involve interactions between institutions such as governments, government agencies, police departments schools, and hospitals. These interactions often consist of interactions between individuals, for example when a political leader gives orders to a police chief. But the interactions between institutions that determine their polices can also depend on whole groups of individuals interacting, for example when a group of politicians from the federal level meet with a group of leaders from the state or provincial level.

Now we can see why systemic racism is more than individual racism because it involves (1) the interactions that are controlled through institutions and (2) the emergent properties of institutions in the form of policies, rules, values, and norms. Overcoming racism, therefore, is more than changing the minds of individuals but also requires institutional modification. Racism cannot be defeated without modifying governments, police departments, schools, and even families. In general, systemic reforms are more than individual reforms because they require changing how people communicate and how they are affected by the policies, values, rules, and norms of institutions, which requires changing the properties of institutions, including not only explicit policies but also implicit values and norms that influence behaviors.

All of these measures can contribute to reinformation by pushing from misinformation to real information.

## **Chapter 8: Misinformation Self-Defense: A Manual Illustrated by the**

### **Ukraine War**

#### **Web Sites for Fact Checking**

Debunker: <https://debunker.club>

Factcheck: <https://www.factcheck.org>

Quackwatch: <https://quackwatch.org>

Skeptic: <https://www.skeptic.com>

Snopes: <https://www.snopes.com>

<https://counterhate.com/>

## **Chapter 9: Reality Rescued**

### **Ancient Anti-Realist Analogies**

The arguments of Seth and Hoffman that I rebutted have ancient predecessors. As chapter 8 describes, I asked ChatGPT “How is Plato's Cave like Zhuangzi's dream of being a butterfly?” and got the following impressive response.

Plato's "Allegory of the Cave" is a philosophical metaphor that appears in Plato's Republic. In the allegory, Plato describes a group of people who have been imprisoned in a cave from birth, and who are only able to see shadows on the cave wall. These shadows are the only reality that the prisoners know, and they believe that the shadows are all there is.

Zhuangzi's "Dream of Being a Butterfly" is a story that appears in the Zhuangzi, a Daoist text attributed to the Chinese philosopher Zhuangzi. In the story, Zhuangzi dreams that he is a butterfly, flying and flitting about freely. When he wakes up, he is not sure whether he is a man who dreamed he was a butterfly, or a butterfly who is now dreaming that he is a man.

Both the Allegory of the Cave and the Dream of Being a Butterfly explore the idea of reality and perception. In the Allegory of the Cave, Plato suggests that our perception of reality is limited and imperfect, and that



there is a higher, more complete form of reality that we cannot fully comprehend. In the Dream of Being a Butterfly, Zhuangzi suggests that reality may be subjective and that our perception of it is shaped by our experiences and circumstances.

Both stories also explore the theme of enlightenment or awakening. In the Allegory of the Cave, the prisoners are said to be "enlightened" when they are able to escape the cave and see the true form of reality. In the Dream of Being a Butterfly, Zhuangzi's dream suggests that enlightenment may involve a realization that our perception of reality is not fixed or certain, but rather is shaped by our individual experiences and circumstances.

ChatGPT and similar new generative AI models require close scrutiny to examine their perils of misinformation.<sup>35</sup>

I think that the allegories of Plato and Zhuangzi are both misleading. Understanding how the brain processes information to perceive the world shows the vacuity of two ancient analogies used to challenge realism. Plato suggested that we are like prisoners in a cave seeing only shadows projected on a wall that they mistake for reality. But a fire shadow could not provide the representational richness that brains get from perceptual interaction with reality. Zhuāngzi dreams of being a butterfly, but a butterfly could not dream of being a philosopher because butterflies only have around a million neurons which is not enough to produce neural representations such as “I am a butterfly dreaming of being a philosopher.”

### **Notes for Supplementary Material**

- 
- <sup>1</sup> *Bots and Beasts*: Thagard 2021a.
- <sup>2</sup> Primitive information: Gleick 2012.
- <sup>3</sup> Probabilities: Shannon 1948, Skyrms 2010.
- <sup>4</sup> Representing and inference: Anderson 2010, Thagard 2005.
- <sup>5</sup> Computational models (samples). Collecting and representing: Eliasmith 2013, Thagard 2019a. Evaluating and transforming: Thagard 2000, 2012. Saving and retrieving: Eliasmith 2013. Sending and receiving: Thagard 2000, Sun 2012.
- <sup>6</sup> Information based on probabilities: Shannon 1948, Dretske 1981, Skyrms 2010.
- <sup>8</sup> Luciano Floridi: Floridi 2013, 2019.
- <sup>10</sup> Psychological information: Miller 1953.
- <sup>11</sup> Psychological information processing: Miller 1956; Newell, Shaw, and Simon 1958; Lindsay and Norman 2013.
- <sup>12</sup> Misinformation processing: Ecker et al. 2022; Gordon et al. 2017; Greiffeneder, Jaffé, Newman, and Schwartz 2021; Lewandowsky, Ecker, and Cook 2017; Lewandowsky et al. 2020, Pennycook and Rand, 2021.
- <sup>13</sup> Illusory truth effect: Hassan and Barber 2021.
- <sup>14</sup> Mere-exposure effect: Montoya, Horton, Vevea, Citkowicz, and Lauber 2017.
- <sup>15</sup> Motivated reasoning effect: Kunda 1990, Hornsey 2020.
- <sup>16</sup> Identity effect. Oyserman and Dawson 2020.
- <sup>17</sup> Confirmation bias: Nickerson 1998.
- <sup>18</sup> Continued influence: Johnson and Siefert 1994, Siefert 2014, Ecker and Antonio 2021.
- <sup>19</sup> Correcting false information: Ecker, Lewandowsky, Swire, and Chang 2011.
- <sup>20</sup> Inoculation: Banas and Rain 2010; Cook, Lewandowsky, and Ecker 2017.
- <sup>21</sup> Nudge intervention: Pennycook, McPhetres, Zhang, Lu, and Rand 2020; Pennycook and Rand 2021, Pennycook et al. 2021.
- <sup>22</sup> Nudge intervention replication: Roozenbeek, Freeman, and van der Linden 2021. Response: Pennycook and Rand forthcoming.
- <sup>23</sup> Backfire effect: Nyhan 2021.
- <sup>24</sup> Institutions: March and Olsen 1989.
- <sup>25</sup> Bayes: Pearl 1988.
- <sup>26</sup> Versus Bayesian approaches: Thagard 2012, 2019c.
- <sup>27</sup> Social values and norms: Thagard 2019b, ch. 5.
- <sup>28</sup> Institutions as rules: Guala 2016, p. xiii.
- <sup>29</sup> Systemic racism: Feagin 2006, Braveman et al. 2022.
- <sup>30</sup> Systemic reform: Chater and Loewenstein ADD.
- <sup>31</sup> CESM: Bunge 2003.
- <sup>32</sup> Social mechanism: Thagard 2019b, p. 51.
- <sup>33</sup> Feagin on racism: Feagin 2006.
- <sup>34</sup> Values and norms: Thagard 2019b, ch. 5.
- <sup>35</sup> ChatGPT perils: <https://www.scientificamerican.com/article/ai-platforms-like-chatgpt-are-easy-to-use-but-also-potentially-dangerous/>.

### References for Supplementary Material

- Albright, Thomas D. "Why Eyewitnesses Fail." *Proceedings of the National Academy of Sciences* 114, no. 30 (Jul 25 2017): 7758-64.
- Anderson, J. R. *Cognitive Psychology and Its Implications*. 7th ed. New York: Worth, 2010.

- 
- Banas, John A., and Stephen A. Rains. "A Meta-Analysis of Research on Inoculation Theory." *Communication Monographs* 77, no. 3 (2010): 281-311.
- Braveman, Paula A, Elaine Arkin, Dwayne Proctor, Tina Kauh, and Nicole Holm. "Systemic and Structural Racism: Definitions, Examples, Health Damages, and Approaches to Dissmantling." *Health Affairs* 41, no. 2 (2022): 171-78.
- Bunge, Mario. *Emergence and Convergence: Qualitative Novelty and the Unity of Knowledge*. Toronto: University of Toronto Press, 2003.
- Butler, Heather A., Christopher P. Dwyer, Michael J. Hogan, Amanda Franco, Silvia F. Rivas, Carlos Saiz, and Leandro S. Almeida. "The Halpern Critical Thinking Assessment and Real-World Outcomes: Cross-National Applications." *Thinking Skills and Creativity* 7, no. 2 (2012): 112-21.
- Chater, Nick, and George Loewenstein. "The I-Frame and the S-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray." *Behavioral and Brain Sciences* 46 (2023): e147.
- Dretske, F. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press, 1981.
- Ecker, U. K. H., and L. M. Antonio. "Can You Believe It? An Investigation into the Impact of Retraction Source Credibility on the Continued Influence Effect." *Memory and Cognition* 49, no. 4 (May 2021): 631-44.
- Ecker, U. K., S. Lewandowsky, B. Swire, and D. Chang. "Correcting False Information in Memory: Manipulating the Strength of Misinformation Encoding and Its Retraction." *Psychonomic Bulletin & Review* 18, no. 3 (Jun 2011): 570-8. <https://doi.org/10.3758/s13428-011-0111-1>.
- Feagin, Joe. *Systemic Racism: A Theory of Oppression*. New York: Routledge, 2013.
- Floridi, Luciano. *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford: Oxford University Press, 2019.
- . *The Philosophy of Information*. Oxford: Oxford University Press, 2013.
- Gleick, James. *The Information: A History, a Theory, a Flood*. New York: Vintage, 2011.
- Gordon, Andrew, Jonathan CW Brooks, Susanne Quadflieg, Ullrich KH Ecker, and Stephan Lewandowsky. "Exploring the Neural Substrates of Misinformation Processing." *Neuropsychologia* 106 (2017): 216-24.
- Greifeneder, Rainer, Mariela Jaffé, Eryn Newman, and Norbert Schwarz. *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation*. London: Routledge, 2021.
- Guala, Francesco. *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton: Princeton University Press, 2016.
- Hassan, Aumyo, and Sarah J Barber. "The Effects of Repetition Frequency on the Illusory Truth Effect." *Cognitive research: principles and implications* 6, no. 1 (2021): 1-12.
- Hilbert, Martin. "Toward a Synthesis of Cognitive Biases: How Noisy Information Processing Can Bias Human Decision Making." *Psychological bulletin* 138, no. 2 (2012): 211.
- Johnson, Hollyn M, and Colleen M Seifert. "Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences." *Journal of experimental psychology: Learning, memory, and cognition* 20, no. 6 (1994): 1420.
- Lewandowsky, Stephan .et al. "The Debunking Handook 2020." (2020). <https://www.climatechangecommunication.org/debunking-handbook-2020/>.

- 
- Lindsay, Peter H, and Donald A Norman. *Human Information Processing: An Introduction to Psychology*. Cambridge, MA: Academic press, 2013.
- March, James G, and Johan P Olsen. "The New Institutionalism: Organizational Factors in Political Life." *American Political Science Review* 78, no. 3 (1983): 734-49.
- Miller, George A. "What Is Information Measurement?". *American Psychologist* 8, no. 1 (1953): 3-11.
- Miller, George A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (1956): 81-97.
- Montoya, R Matthew, Robert S Horton, Jack L Vevea, Martyna Citkowicz, and Elissa A Lauber. "A Re-Examination of the Mere Exposure Effect: The Influence of Repeated Exposure on Recognition, Familiarity, and Liking." *Psychological Bulletin* 143, no. 5 (May 2017): 459-98.
- Newell, A., J. C. Shaw, and H. Simon. "Elements of a Theory of Human Problem Solving." *Psychological Review* 65 (1958): 151-66.
- Nyhan, B. "Why the Backfire Effect Does Not Explain the Durability of Political Misperceptions." *Proceedings of the National Academy of Sciences* 118, no. 15 (Apr 13 2021): 1-7. <https://doi.org/10.1073/pnas.1912440117>.  
<https://www.ncbi.nlm.nih.gov/pubmed/33837144>.
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems*. San Mateo: Morgan Kaufman, 1988.
- Pennycook, G., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand. "Shifting Attention to Accuracy Can Reduce Misinformation Online." *Nature* 592 (Mar 17 2021): 590-95.
- Pennycook, G., and D. G. Rand. "The Psychology of Fake News." *Trends in Cognitive Sciences* 25, no. 5 (May 2021): 388-402.
- Pennycook, Gordon, Jonathon McPhetres, Yunhao Zhang, Jackson G Lu, and David G Rand. "Fighting Covid-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention." *Psychological science* 31, no. 7 (2020): 770-80.
- Pennycook, Gordon, and David G Rand. "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning." *Cognition* 188 (2019): 39-50.
- Pennycook, Gordon, and David G. Rand. "Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation." *Nature Communications* (forthcoming).
- Roozenbeek, Jon, Alexandra LJ Freeman, and Sander van der Linden. "How Accurate Are Accuracy-Nudge Interventions? A Preregistered Direct Replication of Pennycook Et Al.(2020)." *Psychological science* 32, no. 7 (2021): 1169-78.
- Seifert, Colleen M. "The Continued Influence Effect: The Persistence of Misinformation in Memory and Reasoning Following Correction." In *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences*, Ed. D. N. Rapp and J. L G. Braasch, Cambridge, Ma: Mit Press Fix, 39-71, 2014.
- Shannon, Claude E. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27, no. 3 (1948): 379-423.

- 
- Shaw, Amy, Ou Lydia Liu, Lin Gu, Elena Kardonova, Igor Chirikov, Guirong Li, Shangfeng Hu, *et al.* "Thinking Critically About Critical Thinking: Validating the Russian Heighten® Critical Thinking Assessment." *Studies in Higher Education* 45, no. 9 (2019): 1933-48.
- Skyrms, Brian. *Signals: Evolution, Learning, and Information*. Oxford: Oxford University Press, 2010.
- Thagard, P. *Mind: Introduction to Cognitive Science*. 2nd ed. Cambridge, MA: MIT Press, 2005.