We Need a Credible Theory of Gullibility

Joachim I. Krueger

Brown University

Claudia Vogrincic-Haselbacher

University of Graz

Anthony M. Evans

Tilburg University

Correspondence:

Joachim I. Krueger
Department of Cognitive, Linguistic & Psychological Sciences
Brown University, Box 1820
190 Thayer Street

Providence, RI 02912
Phone: (401) 863-2503
E-mail: Joachim_Krueger@Brown.edu

Abstract

Lest the term 'gullibility' remains confined to moralistic judgment, theoretical work is needed to make the concept thick and to ground it in a general psychological space. We approach this project by reviewing instances of gullibility from various perspectives on inductive reasoning (Humean, Bayesian, decision-theoretic). The easy task is to represent gullibility as a special case of heuristic reasoning and predictable irrationality. The difficult task is to embed gullibility in a theory that describes successes and failures of reasoning under lawful conditions. Another difficulty is to stake the concept's boundaries. Where does gullibility end? We use the issue of irrational trust as an example to explore whether a closely related concept may shed light on gullibility proper.

There's an old saying in Tennessee — I know it's in Texas, probably in Tennessee — that says, fool me once, shame on — shame on you. Fool me — you can't get fooled again.
        ~ George W. Bush[1]

George W. Bush may have fooled his fellow Americans many times, but he did not fool everyone all the time. Occasionally, the evidence to the contrary of what he asserted punched through. His and Dick Cheney's claim that Iraq harbored weapons of mass destruction (WMD) at the time the United States was preparing for war was not substantiated – to the point that the absence of evidence amounted to evidence of absence (Hartnett & Stengrim, 2004). Of course, the evidence (or lack thereof) did not move all minds. Even today, some people feel that Saddam Hussein had such weapons somewhere (Hochschild & Einstein, 2015). These believers may point to the fact that Saddam did use WMD on citizens of his own country when unleashing poison gas on the Kurds. If WMD were found, dedicated minds might argue that the CIA had planted these weapons in order to reconcile ordinary American minds with the idea of war. As long as an idea is neither theoretically impossible nor empirically disproven, even David Hume or Thomas Bayes can only ask for skepticism but not outright disbelief.

As a low-stakes example, consider a social media post in which one of the authors (JIK) announced his departure from academia for the corporate world. Some of his friends wished him well, whereas others expressed puzzlement. Responding to a call from his super-ego, he eventually revealed his jocular intent. Humor is a risky form of gullibility play. The humorous guller must trust the gullees to 'get it.' The humorous guller must, in other words, temporarily disable the Gricean norm of truthfulness. By contrast, the serious guller strategically leverages the Gricean norm, thereby debasing its currency. Hence, the issue of gullibility can be framed as a dilemma or as a dialectic. This is the approach we take in this chapter.

---

[1] This quote was retrieved from https://www.goodreads.com/quotes/6436-there-s-an-old-saying-in-tennessee-i-know-it-s

The gullibility game

Parents tell their children – much as pastors, rabbis, schoolteachers, and insurance salespeople tell the rest of us – 'thou shalt not lie!' Deontological imperatives are a culture's way of creating consent at low cost (Baron, 2012). These imperatives ask us not to think and reflect, but to engage the unquestioning, reflexive capacity of mind. This tends to work well enough to hold society together, but, as game theory teaches, being truthful is not a dominating strategy (Binmore, 2007). You will not be better off being truthful irrespective of whether others lie. Because Kant knew this, he declared principled truthfulness moral. Nor is lying a dominating strategy as defection is in the prisoner's dilemma. The assurance game, which is also known as the 'stag hunt' (Rousseau, 1754/1984), is a better model for the gullibility game than the prisoner's dilemma (Krueger, Evans, & Heck, 2017). The assurance game raises the problem of bilateral trust. In a two-person game, both players prefer to cooperate if the other cooperates. Fearing that the other might defect, however, each player prefers unilateral defection to unilateral cooperation. Why might a player fear that the other might defect? Because this player might fear that the other player might fear that this first player might have that fear, and so on. The point is that the assurance game has no dominating strategy. Setting aside context effects, communication, or privileged knowledge, a tit-for-tat ("equivalent retaliation") heuristic – introduced by Anatol Rapoport and *mis en scène* by Robert Axelrod (1984) is a prudent and pragmatic strategy for these games. Speak the truth unless you are lied to.

Any version of game theory demands that there be clear knowledge of whether an assertion is truthful once it is uttered. Often we don't know which is the case, and we must

apply personal standards of how much credence we will grant a claim when the claimant's intentions are veiled.

Gullibus dialecticus

A dialectic is a potential intellectual impasse, a fanciful way of saying 'it's complicated,' but with a hope of resolution. If we cannot trust all claims all the time for fear of being gulled, how do we know the difference? If gullibility is the thesis and cynicism is the antithesis, how do we find a synthesis? To say 'believe the credible' is to beg the question of how we know what is credible. Psychological science is flush with theories of animal learning (Church, 1963), deep learning (LeCun, Bengio, & Hinton, 2015), very deep learning (Simonyan & Zisserman, 2014), and machine learning (Rasmussen, 2004). All these theories assume the existence of honest feedback. The signal may come with noise, and perception and memory may be imperfect, but the feedback is rarely allowed to be strategic – unless experimenters lie to their participants[2].

Epistemology and induction

Being skeptical during the generally optimistic era of the enlightenment, Hume allowed the possibility of gulling. In his essay on miracles, he maintained that extraordinary claims require extraordinary evidence (Hume, 1748/1959). With this definition of a miracle, he assigned a low probability of it occurring. For a claim of a miracle to be accepted, the probability of the claim being fake – by mistake or by design – must be lower still. Hume advised that we should not accept such claims lightly; the fact that beliefs in miracles are not uncommon suggests that Hume's rule has normative force (Hájek, 1995).

The same holds for Bayes' rule, although Bayes was not concern with lying. Bayes' equation, which shows the relations among conditional and unconditional probabilities, is normative in the mathematical sense. Thou shalt not violate Bayes' theorem lest thou be

---

[2] Experimental social psychology traditionally relies on deception, and thus on the participants' gullibility.

incoherent. In psychological science, Bayes' theorem is often accepted as normative (Edwards, Lindman, & Savage, 1963). Thou shalt think in the Bayesian way lest thou be irrational (Dawes, 1988). The psychological idea expressed by Bayes' theorem is seemingly beyond reproach. Belief – broadly defined as a person's subjective sense of the likelihood of an assertion being true – should be sufficiently determined by prior belief and evidence – in multiplicative fashion. The strength of the prior belief may be expressed as a likelihood or a probability, p(B), and the strength of the evidence is expressed by the ratio of the likelihood of the evidence assuming that the belief in question is true, p(D|B), over the cumulative likelihood of the evidence under all conceivable beliefs, p(D). When only two mutually exclusive and exhaustive beliefs are in play ('The aliens either landed or they did not'), the latter term reduces to the sum of the likelihood of the evidence assuming that the belief is true plus the likelihood of the evidence assuming that the belief is false.

Bayes' theorem and the human judgment it informs work well when the numbers are on the table, either as probabilities or as frequencies (Gigerenzer & Hoffrage, 1995). Basic rationality is assumed in that the human judges are not expected to bicker over the numbers. They only need to do know how to divide and multiply. Although Bayesian theories model subjectivity, they demand consensus on what the data are. Subjectivity is placed in the term p(B); it may vary over judges. The evidence, or the data, D, is objective. Thus, the term p(D|B) is objective. The assumption that humans honor the objectivity of p(D|B) shows the optimism of the enlightenment. Yet, human cognition and perception are subject to motivations and cognitive constraints, so much so that a value of, say, .05 for p(D|B) means different things to different people.

A third normative approach is decision theoretic (Swets, Dawes, & Monahan, 2000). After Hume and Bayes, we meet Pascal (Hájek, 2003; Krueger, 2011a). Here too, we find a family of models and techniques, but the basic idea is simple. Decisions bisect reality so that the decision to endorse belief B amounts to a true positive, or a "Hit," if that belief is true. If this belief is not true, the result is a false-positive error. If the decision is to reject B, but B is true, the result is a false-negative error, or "Miss." If B is rejected and it is false, the result is a 'Correct Rejection.' A simple epistemic and non-utilitarian goal is to maximize accuracy by reducing the overall proportion of erroneous decisions. If, however, the relative values assigned to the four possible outcomes vary, so should the decision threshold or bias to endorse the belief. Following this logic, Pascal wagered to believe in God, considering it more painful to reject an existing god than to believe in a non-existing one. Pascal was a belief-liberal.

By contrast, conventional significance testing is belief-conservative as it is biased against beliefs in the non-existent. Here, the term p(D|B) refers to the p value associated with the test statistic. Only if $p < .05$, the researcher rejects the idea of nothing- or nullness, tentatively inferring that there is systematic variance (an effect!) (Krueger & Heck, 2017). Significance testing is the researcher's way of checking gullibility. There is intense debate as to whether the protections against being gulled by the data (or naïve or nefarious investigators) are sufficient, which reinforces the idea that most people (lay- and academic alike) tend to have a greater fear of believing something that is not there than of missing something that is (Ioannides, 2014). Perhaps these fears are so strong because the evidence – mostly obtained with significance testing – suggests that people hold a far greater portfolio of beliefs than they should.

The common subtext to these induction methods, be they Humean, Bayesian, and Fisherian (excepting Pascal himself), is that lay epistemology is too liberal, thereby

underwriting gullibility. Formal methods of induction must teach the human observer to stick to the evidence. These formal methods have had a greater impact on scientific work than on everyday thinking. Ordinary thinking departs from formal epistemologies in two seemingly contradictory ways. One way is simple gullibility, which is the unreflective willingness to believe a claim. The other way is conditional gullibility, which is the tendency to assimilate evidence to pre-existing beliefs. Let us consider the two in turn.

The power of the given stimulus

The only thing that matters is what's on your mind – not what's in it. ~ Amos Tversky in conversation with Victor Shamas (Shamas, 2018, p. 162)

Following Spinoza, Gilbert (1991) proposed that comprehending a claim and believing it are initially one and the same. To comprehend X, we must first believe X, if only for a moment. Middle school graffiteurs who write "*Wer dies liest ist doof*" (He – or she – who reads this is dumb) understand this. Doubt, skepticism, and eventual disbelief require mental work and thus time. The claim X must be situated within a mental network of related and relevant propositions, and the coherence of this network must be checked. If adding X to the network of beliefs reduces the network's coherence, X may be tagged as being negated. Psychologically, this tagging amounts to a verdict of 'not true' rather than 'false.' Gilbert's theory is not statistical, but perceptual and cognitive. It treats gullibility as the result of the mental system's default operation, which can only be removed after careful consideration of contrary evidence.

Gilbert's (1991) dual-process approach captures only part of the general decision-theoretic model. As his approach focuses on the role of reflective cognition for the identification and the removal of false-positive errors, it limits good thinking to the transformation of false positive beliefs into correct rejections. False negatives remain unknown, and therefore, any assessment of accuracy as the statistical association between belief (vs. unbelief) and reality

(true vs. false) undefined. Gilbert's (1991) theory of belief was an early example of a dual-process theory, which has swept psychology since. Many of the heuristics proposed in Kahneman and Tversky's path-breaking work came to be reinterpreted in dual-process or dual-systems terms (Kahneman, 2011; but see Dawes, 1976, for a early dissent). A latecomer to dual systems, Kahneman suggested that fast, intuitive thinking is perceptual in nature, whereas slow, deliberative thinking represents what is ordinarily considered reasoning. Errors, and irrationality more generally, occur when the intuitive system generates a wrong response and when the deliberative system fails to correct it.

Viewed this way, gullibility flourishes where heuristic reasoning reigns. Anchoring is the effect of an initial stimulus on eventual judgment. The effect is dramatic, as it can occur even when the anchor is extreme or coming from randomness (Frederick & Mochon, 2012; Tversky & Kahneman, 1974). Gullers can use anchoring to induce gullees to accept beliefs that are more extreme than they would be or should be without the anchor. Deliberative thinking works to mitigate the anchor's effect, but insufficiently so (Epley & Gilovich, 2006). Representative thinking is equally well documented. Gullers can induce false beliefs by using evocative imagery, which leads to false judgments if these images contradict statistical considerations. Magicians and charlatans are fond of providing audiences with visual, experiential demonstrations, seeking to disable slow thinking about coherence and probability. What Kahneman and Tversky (1982) called the simulation heuristic capitalizes on the power of story and narrative. Humans are gulled by good stories, that is, stories that have a dramatic arc, reveal causal relations, and tell about the reasons and intentions of human agents (Pennington & Hastie, 1993; Schank & Abelson, 1995). Narrative cognition bestows many benefits, not least among them being boosted memory (Bartlett, 1932/1997). Again, however, gullers can exploit

the human readiness for the narrative by weaving deceptive stories that satisfy these criteria but that happen to be utterly false. Belief is beggared only if one cares to look.

      If heuristics threaten rationality and foster gullibility, we begin to see the connection between the two. Both ought to be overcome by slow and careful thinking, but often are not. Many heuristics – with the exception of the story heuristic – may be bundled due to a shared underlying dynamic: the tyranny of the salient stimulus (Dawes, 1988). People are liable to be gullible if they fail to go beyond what is in front of them (Posner, 1973). Uncorrected, their judgments show focalism and nonregressiveness (Fiedler & Krueger, 2012; see also section epigraph). Social and cognitive psychology have produced a wealth of evidence for a variety of processes that can be subsumed under this rubric. To name a few; stimuli become more salient (and thus gulling) if they are novel, surprising, or familiar (Pennycook & Rand, 2017), if they are vividly imagined or causally explained (Gregory, Cialdini, & Carpenter, 1982), or if people fail to question them (mindlessness; Langer, Blank, & Chanowitz, 1978; see also Forer, 1949). Through various associative processes, salient stimuli affect judgment and inference more broadly. Salient stimuli become emotionally significant through evaluative conditioning (Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012) and diagnostic (or pseudo-diagnostic) of other dimensions through statistical contingency (Fiedler, Freytag, & Meiser, 2009; Rothbart, 2015/1981).

      Whatever makes a stimulus salient and focal is not necessarily indicative of truth, although there may be probabilistic relationships. If a correction is needed, how can it be achieved? This is a problem that still awaits an elegant solution. To illustrate , consider the anchoring heuristic. Kahneman, Tversky, and the investigators they inspired, offer no normative model of adjustment. The error arising from the use of this heuristic is seen in the post-

adjustment difference in the estimates made by those starting with a high anchor and those starting with a low one. Even if this difference were closed, the result might be wildly inaccurate (Krueger, Freestone, & MacInnis, 2013).

This limitation highlights a shortcoming of the dual-systems approach, namely the lack of an explicit decision-theoretic framework. There is a built-in neglect of false-negative errors, that is, no provision is made for cases in which people would have fared better had they used an intuition-based heuristic (Fiedler, Kutzner, & Krueger, 2012). Research in the fast-and-frugal-heuristics framework addresses this issue by abandoning the intuition-plus-correction template (Gigerenzer & Gaissmeier, 2011). Instead, this framework asks questions of ecological validity to distinguish the conditions under which intuitive responses work well from the conditions under which they do not. From this perspective, there is no generic issue of gullibility; instead there is a readiness to believe that may be modulated by contextual conditions, with positive or negative consequences depending respectively on the match or the mismatch of the person's psychological capacities with environmental structure.

Both, the heuristics-and-biases approach and the fast-and-frugal-heuristics approach leave important questions unanswered. Whereas the former often fails to advise human judges just how much to correct their intuitions, the latter offers little assistance in how to choose the best heuristic in a given setting (Rieskamp & Otto, 2006). The human judge has two options, which have only begun to receive research attention. The first option is to turn to processes that stimulate creativity when formal models are mute on how to go beyond the focal stimulus. Turning to a mindset of foraging, the judge can open up associate networks (Baror & Bar, 2016; Colzato, Ozturk, & Hommel, 2012). This tactic may not be sufficient to restore rationality, but it can help prepare the ground. Mental foraging has the potential to dilute or transform claims that

might otherwise gull the person. The second option is to strategically choose ignorance over information, when there is reason to believe that that information is tainted or gulling (Hertwig & Engel, 2016). *Homo ignorans* cannot be *Homo credulus*.

A core assumption of belief-correction models is that the person must want to correct the biasing effect of the focal stimulus (Evans & Stanovich, 2013; but see Krueger, 2012a; Kruglanski, 2013). Sometimes, people just don't want to. Times of distress, uncertainty, and despair are fertile ground for the unchecked growth of belief and superstition (Keinan, 2002). Practices of questionable scientific credibility enjoy popularity when other options have been exhausted (e.g., homeopathy, dowsing, use of charms). Practitioners and their clients alike use the heuristic 'It can't hurt,' and shift the burden of proof to the skeptics (Vyse, 1997/2014). Practitioners can become self-gullers, gathering enough positive evidence to convince themselves that the practice works (Hyman, 1981). Focusing on the co-occurrences of the focal stimulus ("the practice") and a desired outcome ("success"), pseudo-contingencies (Fiedler et al., 2009) and illusory correlations (Dawes, 1989) take in the unwary. Self-gulling becomes a case of self-enhancement and overconfidence (Heck & Krueger, 2015; Moore & Healy, 2008).

The return of the prior belief

From a Bayesian perspective, the power of the focal stimulus over the belief *du jour* shows the power of diagnostic or pseudo-diagnostic information over judgment, to the neglect and detriment of prior belief. Tversky and Kahneman (1974) themselves famously asserted that focalism is so strong that it cannot be understood as a poor form of Bayesian thinking but that it is something entirely different. Yet, closing the book of reasoning at this juncture would itself be an example of focalism. There is more.

The idea of gullibility as a losing battle against the focal stimulus misses an important psychological element. It is not the case that people believe everything and anything at any time. Gullibility can be highly conditional. Some people categorize claims into the credible and the incredible *a priori*. As an extreme example, consider popular conspiracy theories. There are a number of ways in which such theories can be developed and maintained. The simplest heuristic is to divide claims into two categories: conventional and subversive. A person applying this distinction may believe any claim falling in the latter class and reject all others. Epistemology, evidence, and truth aside, this tactic can yield great motivational benefits. Subscribers may feel 'clued in,' privileged, and being members of a select few who have peered behind the veil (Krueger, 2010). An acceptance of all breaks with convention can lead to grotesque contradictions. With this strategy, reports of moon landings will be considered fake, while reports of aliens brought to earth by Apollo 20 (twenty!) ring true. More poignantly, a person with this contrarian mindset may claim that climate change is a hoax, while believing that the government manipulates the weather to scare the populace (Lewandowsky, Oberauer, Gignac, 2013). The common denominator is the idea that the government and traditional news outlets lie by default (Pennycook, Cannon, & Rand, 2017). Ironically then, a conspiracy theorist of this type can cultivate a self-image of ingullibility. It is the ordinary people that live behind a veil of ignorance and deception because of their gullibility. Again, gullibility begets self-enhancement. At the extremity, a conspiracy theorist sees no contradiction in the belief that the government practices mind control, but that his or her own mind has not thus been controlled.

An extreme variant of conditional gullibility is gaslighting, further amplified in self-gaslighting. To gaslight someone is to insist on a falsehood with such conviction that the victim begins to doubt his or her sanity (Abramson, 2014). Some claims of miracles my require

gaslighting to get any traction. A communicator may boldly claim that Nelson Mandela died in prison until some listeners begin to consider this as a possibility, if only in a parallel universe (Krueger, 2016). Tertullian raised gaslighting to a principle of belief when declaring *credo quia absurdum*, "I believe it because it is absurd" (Bühler, 2008). Variants of gaslighting can be observed on the political scene on a daily basis ("The crowd at my rally was the largest ever!").

A more sophisticated gullibility heuristic is the assumption that there is a higher order or logic to things, which only the discerning person can appreciate. In esoteric circles, the clichés that 'the universe does not make mistakes,' that there is a 'cosmic balance' to things, or that 'everything is ultimately for the best' have great power to gull the mind and fill it with nonsense. Instead of setting up conditions for disbelief, this heuristic 'liberates' the person from such conditions if they were suggested. Consider the case of Robert Betz, a self-help promoter and impresario. In a workshop attended by JIK, Mr. Betz asserted that disease is the way of the universe to keep score (Krueger, 2008). Leaving open the question of whether misbehavior eventually begets disease, Betz claimed that disease is always caused by misbehavior. The universe let's us know that this is so by making the disease representative of the offense. Parkinson's disease, by this logic, both reveals and punishes the afflicted person's earlier quests for control. When asked by a woman in the audience how he explains birth defects, Betz responded with the only argument left: the person's misbehavior lay in a past life. When asked by JIK how he knew this to be so, Betz put his hand on his chest and declared that he knew this in his heart. Belief had become untethered from reality. No other challenges were made.

While Mr. Betz holds a master's degree in marketing, Professor Justin Barrett has contributed to the empirical literature on child development, and in particular the genesis of the belief in god (Barrett, 2012). Yet, Barrett puts belief before epistemology (Krueger, 2011b;

2012b). His proof of the existence of god goes like this: 'If god exists, we may assume that He created man in such a way that man is prepared to believe in god. Now that I and many others believe in god, I conclude that god exists.' This is a case of the reverse-inference fallacy (Krueger, 2017). Effectively, Barrett's argument uses his conclusion as evidence for its own truth.

Gullibility, in this world, can be self-sustaining. Illogic and irrationality provide multiple psychological tools, but the greater context is social. False and destructive beliefs survive in part when the public, even if enlightened, is excessively polite. It is generally difficult to challenge bad ideas, especially in public (Asch, 1956; Krueger & Massey, 2009). Conformity and acquiescence are powerful forces, evolved to support social cohesion and peace. To confront lies, deception, and bad ideas, and do so assertively, requires skill and strength of mind (see von Horváth, 1937/2017, for a literary treatment of this issue). Professional gullers know what they can get away with, and it tends to be a lot. One particularly potent device is to signal to the audience that the (false) claims made are part of or are embedded in deep values. The deepest values are treated as sacred possessions (Tetlock, 2003), that is, with the highest respect. When, however, these claims are false or exploitative, resistance is needed, not acquiescence.

As an example of personal failure in this regard, consider another visitation of the conspiracy scene (Krueger, 2015ab). In this instance, one of us (JIK), attended a 'congress' of Russian 'scientists' billed as 'Mindfully into the future.' One of the presenters, a certain Prof. Dr. Sergej Sall, asked who benefits from the antagonism between the Western and the Islamic worlds. The answer would appear to be obvious to those with an adequately prepared mind, but Sall added, for good measure, an oracular sign right out of the heuristics-and-biases toolbox. Hidden truths reveal themselves in surface similarities. Hence, Sall presented the claim –

without commentary – that ISIS is an acronym for Israel Secret Intelligence Service. The skeptical and rational mind might wonder why, if this service were so secret, would it leave such an obvious hint. There was no open resistance from the audience (including JIK), only polite applause (not from JIK).

Evidence checked

"Evidence checked" is a *double entendre*. People exposed to claims can check the relevant evidence to regulate their beliefs, or they can 'check' the evidence as they might check (i.e., resist) an enemy's advance. At one pole of this continuum, the pure form of enlightenment thinking demands the admittance and fair evaluation of all relevant evidence, and a principled updating of one's beliefs, perhaps in Bayesian fashion. At the other pole, there is outright neglect or rejection of the evidence, as described in the previous section. Most human psychology plays out in the middle ground where evidence is entertained, but selectively so.

Beginning with Festinger's (1957) theory of cognitive dissonance, the strategic selectivity of the social mind has been a stock presence in research on attitudes and attitude change as well as judgment and decision-making more general (Fischer, 2011; Frey, 1986). Whether this selective sampling of information (Fiedler & Juslin, 2006) and its biased assimilation (Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998; Lord, Ross, & Lepper, 1979) is motivated or a built-in feature of the cognitive process is not of concern here. Rather, we note that selective processing provides a self-reinforcing set of mechanisms that have the potential of locking in false beliefs originating from focal stimuli. In other words, selective processing exacerbates gullibility.

Traditionally, social psychology has been concerned with the effects of (potentially insincere) individual communicators (Cialdini, 2016) or the effects of institutionalized

propaganda (Lewandowsky, Stritzke, Freund, Oberauer, & Krueger, 2013). In today's digital

world, social media play a big role in shaping beliefs on just about any topic. Here, both the

individual's biased choices and the platform's tailored algorithms converge on highly selective

exposure to news and stories (Bakshy, Messing, & Adamic, 2015). Users find themselves in

"echo chambers" and "filter bubbles" that amplify rather than test or modulate initial beliefs.

Lack of awareness of this skew, besides a lack of will or capacity to correct it, raises concerns

about gullibility to a higher level (Knobloch-Westerwick, Mothes, & Polavin, 2017).

Susceptibility to "fake news" has become the postmodern face of gullibility *par

excellence*. Defined as news items that are "intentionally and verifiably false" (Allcott &

Gentzkow, 2017, p.4), or that "contradict the best available evidence" (Flynn, Nyhan, &

Reifler, 2017, p. 2), the term fake news has been popularized by politicians who themselves

hold dubious records of truthfulness. Because of its pervasiveness and resistance to

correction, fake news compromises a democracy's functioning (Pennycook et al., 2017).

Fake news contaminates public discourse on the economy (Bartels, 2002), foreign policy

(Kull, Ramsay, & Lewis, 2003), gun control (Aronow & Miller, 2016), climate change

(McCright & Dunlap, 2011), vaccination (Freed, Clark, Butchart, Singer, & Davis, 2010),

and genetically modified food (Gaskell et al., 2004). At a time when anything presented or

represented on the internet qualifies as "information," psychological limitations to the

person's ability or willingness to correct false beliefs is not the only or even the central

concern. Even an educated and willing mind is easily gulled because it does not even know

which items require correction or dismissal. "Bullshit," to use Harry Frankfurt's (2005)

technical term, is as hard to detect as it is to correct. Contemporary society is called upon to

find new ways of creating and maintaining trusted authorities for the dissemination of

credible and evidence-based information in the public interest.

Gullibility without a guller: the case of irrational trust

It is hard to imagine gullibility without an agent doing the gulling. The Hebrew creation myth features a serpentine guller praying on human vanity, thereby introducing gullibility as part of the original sin. Yet, we have seen examples of self-gulling, which plays out daily on millions of computer and smartphone screens. We have argued that a wholesale condemnation of susceptibility to social influence is no answer to this situation. Humans need to rely on the testimony of others to some extent; they cannot function without trust. Like gullibility more generally, interpersonal trust presents a dilemma. Without trust, life is poor; with trust, it might be a disaster (Krueger & Evans, 2013). Collectively, humans flourish if they trust one another (Johnson & Mislin, 2011), but those who are being trusted have an incentive to defect, especially if the interaction is short-lived.

In the experimental trust game, the issue of gullibility manifests as naivité. The game does not require a gulling agent. Players can simply decide if they wish to transfer money to another person, knowing that the transfer creates value (typically, the transferred amount is tripled), and knowing that the other person may or may not reward an act of trust by respectively returning part of the new wealth or by keeping it all (Evans & Krueger, 2009). If the problem of gullibility is that people believe too much, the problem of naivité is that people might trust too much. Conventional game theory asserts that any act of trust is irrational because the trustee has no real (i.e., material) incentive to give back. This understanding of rational choice is too restrictive, however. The break-even point is a more realistic criterion. If people invest more than they receive back, their trust seems exaggerated.

Some have argued that trust is inflated because people perceive and enact it as a default of cooperation (Dunning, Anderson, Schlösser, Ehlebracht, and Fetchenhauer, 2014; Rand, 2016). They think it is the socially correct thing to do, and pay the price in lost returns (Krueger, Massey, & DiDonato, 2008). On this view, trusting is a social norm, on a par with the norm of reciprocity (Gouldner, 1960). This hypothesis is only partly valid. People are willing to punish those who fail to reciprocate trust, but refuse to punish those who fail to trust in the first place (Bicchieri, Xiao, & Muldoon, 2011). It is indeed implausible that parents and padres enjoin their charges to trust blindly. They rather seek to instill awareness that trust is a dilemma that can only be mitigated by figuring out whom and when to trust.

The trust dilemma is hard, and people have no more than bounded success solving it (Evans & Krueger, 2016). Although they are sensitive to their own potential gains and losses, people do not pay much attention to the other person's (the trustee's) incentives. These incentives predict behavioral trustworthiness well, but those who trust – or decide not to – neglect this source of information. As a result, their decisions are insufficiently regressive. Both, strong trust and strong distrust, should be tempered (Evans & Krueger, 2017). People are not only too impressed by their own potential payoffs as focal stimuli (Evans & Krueger, 2011), they also rely too much on the physical attractiveness of others when deciding whether to trust (Bonnefon, Hopfensitz, & De Neys, 2013; Olivola, Funk, & Todorov, 2014). Instead of being a generic default, trust is cue-dependent; and some of the cues are bad. But people learn. Trust decisions mature during childhood (Evans, Athenstaedt, & Krueger, 2013) and they are correlated with intelligence and reflective reasoning (Corgnet, Espín, Hernán-González, Kujal, & Rassenti, 2016).

Review and outlook

In this chapter, we considered elements of a credible theory of gullibility without proposing such a theory. The challenges to such an endeavor are considerable. To be unique, a credible theory of gullibility would have to treat gullibility as a psychological phenomenon that cannot be reduced to general principles of learning, persuasion, and belief change. Such a theory would have to ensure that usage of the term gullibility is not judgmental, moralistic, or subject to outcome bias (Baron & Hershey, 1988; Heck & Krueger, 2016). We submit that a fruitful way of theorizing about gullibility is within the context of broader perspectives on inductive reasoning (Humean, Bayesian, Pascalian, as well as dual-processes approaches). Another theoretical challenge is to account for special cases of self-gulling or of being gulled in the absence of communication. We considered the phenomenon of irrational trust to explore the border region between gullibility proper and a related but overlapping concept.

References

Abramson, K. (2014). Turning up the lights on gaslighting. *Philosophical Perspectives, 28*, 1–30. doi:10.1111/phpe.12046

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *NBER Working Paper Series*, *Working Paper 23089*.

Aronow, P. M., & Miller, B. T. (2016). Policy misperceptions and support for gun control legislation. *The Lancet*, *387*(10015), 223. https://doi.org/10.1016/S0140-6736(16)00042-8

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monograph, 70*(9, whole number 416).

Axelrod, R. (1984). *The Evolution of Cooperation.* New York, NY: Basic Books.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Baron, J. (2012). Where do nonutilitarian moral rules come from? In J. I. Krueger (ed.), Social judgment and decision making (pp. 261-278). New York, NY: Psychology Press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*, 569-579. http://dx.doi.org/10.1037/0022-3514.54.4.569

Baror, S., & Bar, M. (2016). Associative activation and its relation to exploration and exploitation in the brain. *Psychological Science, 27*, 776-789. DOI: 10.1177/0956797616634487

Barrett, J. L. (2012). *Born believers: The science of children's religious belief*. New York, NY: Free Press.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*, 117–150.

Bartlett, F. C. (1932/1997). *Remembering: A study in experimental and social psychology*. Cambridge, U: Cambridge University Press.

Bicchieri, C., Xiao, E., & Muldoon, R. (2011). Trustworthiness is a social norm, but trusting is not. P*olitics, Philosophy & Economics, 10,* 170 – 187. https://doi.org/10.1177/1470594X10387260

Binmore, K. (2007). *Game theory: A very short introduction*. Oxford, UK: Oxford University Press.

Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General, 142*, 143-150. http://dx.doi.org/10.1037/a0028930

Bühler, P. (2008). Tertullian: The teacher of the credo quia absurdum. In J. Stewart (ed.). *Kierkegaard research: Sources, reception and resources, Vol. 4: Kierkegaard and the patristic and medieval traditions* (pp. 131-138). Aldershot, UK: Ashgate.

Church, R. M. (1963). The varied effects of punishment on behavior. *Psychological Review, 70*, 369-402. http://dx.doi.org/10.1037/h0046499

Cialdini, R. (2016). *Pre-suasion: A revolutionary way to influence and persuade*. New York, NY: Simon & Schuster.

Colzato, L. S., Ozturk A., & Hommel B. (2012). Meditate to create: the impact of focused-attention and open-monitoring training on convergent and divergent thinking. *Frontiers in Psychology, 3*, 116. doi: 10.3389/fpsyg.2012.00116

Corgnet, B., Espín, A. M., Hernán-González, R., Kujal, P. & Rassenti, S. (2016) To trust or not to trust: Cognitive reflection in trust games. *Journal of Behavioral and Experimental Economics, 64*, 20-27. https://doi.org/10.1016/j.socec.2015.09.008

Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 3-12). Oxford, England: Lawrence Erlbaum.

Dawes, R. M. (1988). *Rational choice in an uncertain world*. San Diego, CA: Harcourt Brace Jovanovich.

Dawes, R. M. (1989). Experience and validity of clinical judgment: The illusory correlation. *Behavioral Sciences & The Law, 7*, 457-467. DOI: 10.1002/bsl.2370070404

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, *75*, 53–69. https://doi.org/10.1037/0022-3514.75.1.53

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology*, 107, 122–141.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*, 193-242. http://dx.doi.org/10.1037/h0044139

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17*, 311-318. https://doi.org/10.1111/j.1467-9280.2006.01704.x

Evans, A. M., Athenstaedt, U., & Krueger, J. I. (2013). The development of trust and altruism during childhood. *Journal of Economic Psychology, 36,* 82-95. doi.org/10.1016/j.joep.2013.02.010

Evans, A. M., & Krueger, J. I. (2009). The psychology (and economics) of trust. *Social and Personality Psychology Compass: Intrapersonal Processes, 3*, 1003-1017. doi:10.1111/j.1751-9004.2009.00232.x

Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk taking and expectation of reciprocity. *Journal of Experimental Social Psychology*, *47*, 171-177. doi:10.1016/j.jesp.2010.08.007

Evans, A. M., & Krueger, J. I. (2016). Bounded prospection in dilemmas of trust and reciprocity. *Review of General Psychology, 20,* 17-28. doi.org/10.1037/gpr0000063

Evans, A. M., & Krueger, J. K. (2017). Ambiguity and expectation-neglect in dilemmas of trust. *Judgment & Decision Making, 12,* 584-595. http://journal.sjdm.org/17/17131/jdm17131.pdf

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives in Psychological Science, 8*, 223–241. https://doi.org/10.1177/1745691612460685

Festinger, L. (1957). A *theory of cognitive dissonance*. Evanston, Ill: Row Peterson, 1957.

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review, 116*, 187-206.

Fiedler, K., & Juslin, P. (2006). *Information sampling and adaptive cognition*. New York, NY: Cambridge University Press.

Fiedler, K., & Krueger, J. I. (2012). More than an artifact: Regression as a theoretical construct. In J. I. Krueger (Ed.). *Social judgment and decision-making* (pp. 171-189). New York, NY: Psychology Press.

Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α-control to validity proper: Problems with a short-sighted false-positive debate. *Perspectives on Psychological Science, 7*, 661-669. doi: 10.1177/1745691612462587

Fischer, P. (2011). Selective exposure, decision uncertainty, and cognitive economy: A new theoretical perspective on confirmatory information search. *Social and Personality Psychology Compass*, *5*, 751–762. https://doi.org/10.1111/j.1751-9004.2011.00386.x

Forer, B. R. (1949). The fallacy of personal validation: A class-room demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44,* 118-123.

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, *38*, 127–150.

Frankfurt, H. G. (2005). *On bullshit*. Princeton, NJ: Princeton University Press.

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General, 141*, 124-133. doi: 10.1037/a0024006

Freed, G. L., Clark, S. J., Butchart, A. T., Singer, D. C., & Davis, M. M. (2010). Parental vaccine safety concerns in 2009. *Pediatrics*, *125*, 654–659. https://doi.org/10.1542/peds.2009-1962

Frey, D. (1986). Recent research on selective exposure to information. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, 41–80). New York, NY: Academic, Press.

Gaskell, G., Allum, N., Wagner, W., Kronberger, N., Torgersen, H., Hampel, J., & Bardes, J. (2004). GM foods and the misperception of risk perception. *Risk analysis: An official publication of the Society for Risk Analysis*, *24*, 185–194. https://doi.org/10.1111/j.0272-4332.2004.00421.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482. doi:10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

Gilbert, D. (1991). How mental systems believe. *American Psychologist, 46,* 107-119. http://dx.doi.org/10.1037/0003-066X.46.2.107

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*, 161-178. DOI: 10.2307/2092623

Gregory, W. L., Cialdini, R. B., Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so?" *Journal of Personality and Social Psychology, 43*, 89-99.

Hájek, A. (1995). In defense of Hume's balancing of probabilities in the miracles argument. *Southwest Philosophy Review, 11*, 111-118. DOI: 10.5840/swphilreview199511111

Hájek, A. (2003). Waging war on Pascal's wager. *The Philosophical Review, 112*, 27-56.

Hartnett, S. J., & Stengrim, L. A. (2004). "The whole operation of deception": Reconstructing President Bush's rhetoric of weapons of mass destruction. *Cultural Studies: Critical Methodologies, 4*, 152-197. https://doi.org/10.1177/1532708603262787

Heck, P. R., & Krueger, J. I. (2015). Self-enhancement diminished. *Journal of Experimental Psychology: General, 144,* 1003-1020. http://dx.doi.org/10.1037/xge0000105

Heck, P. R., & Krueger, J. I. (2016). Social perception of self-enhancement bias and error. *Social Psychology, 47,* 327-339. DOI: 10.1027/1864-9335/a000287

Hertwig, R., & Engel, C. (2016). *Homo ignorans*: Deliberately choosing not to know. *Perspectives on Psychological Science, 11*, 359-372. https://doi.org/10.1177/1745691616635594

Hochschild, J. L., & Einstein, K. L. (2015). Do facts matter? Information and misinformation in American politics. *Political Science Quarterly, 130*, 585-624. DOI: 10.1002/polq.12398

Hume, D. (1748/1959). *Enquiry concerning human understanding*. New York, NY: Dover.

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141*, 539-557. DOI: 10.1037/a0026477

Hyman, R. (1981). The psychic reading. *Annals of the New York Academy of Sciences, 364*, 169-181. DOI: 10.1111/j.1749-6632.1981.tb34471.x

Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Med* 11(10): e1001747. https://doi.org/10.1371/journal.pmed.1001747

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Analysis, 32*, 865-889. https://doi.org/10.1016/j.joep.2011.05.007

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press.

Keinan, G. (2002). The effects of stress and desire for control on superstitious behavior. *Personality and Social Psychology Bulletin, 28*, 102-108. https://doi.org/10.1177/0146167202281009

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2017). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, *7*, 1-21. https://doi.org/10.1177/0093650217719596

Krueger, J. I. (2008). Report on a self-help workshop. *Psychology Today Online*. https://www.psychologytoday.com/blog/one-among-many/200812/report-self-help-workshop Posted December 29, 2008, retrieved February 28, 2018

Krueger, J. I. (2010). Die "Grosse Verschwörungstheorie" aus psychologischer Sicht [The "Great Conspiracy Theory" from a psychological perspective]. *Zeitschrift für Anomalistik, 10*, 6-16.

Krueger, J. I. (2011a). Don't bet on Pascal's wager. *Psychology Today Online*. https://www.psychologytoday.com/blog/one-among-many/201112/don-t-bet-pascal-s-wager Posted December 17, 2011, retrieved February 28, 2018.

Krueger, J. I. (2011b). Belief is not evidence. Psychology Today Online. https://www.psychologytoday.com/blog/one-among-many/201112/belief-is-not-evidence
Posted December 25, 2011, retrieved March 3, 2018.

Krueger, J. I. (2012a). The (ir)rationality project in social psychology: A review and assessment. In J. I. Krueger (Ed.). *Social judgment and decision-making* (pp. 59-75). New York, NY: Psychology Press.

Krueger, J. I. (2012b). Barrett's banalogies. *Psychology Today Online*. https://www.psychologytoday.com/blog/one-among-many/201204/barretts-banalogies Posted April 18, 2012, retrieved March 3, 2018.

Krueger, J. I. (2015a). Gullibility gulag. *Psychology Today Online*. https://www.psychologytoday.com/blog/one-among-many/201511/gullibility-gulag

Krueger, J. I. (2015b). Gulliber's travails. *Psychology Today Online*. https://www.psychologytoday.com/blog/one-among-many/201512/gullibers-travails.
Posted December 19, 2015, retrieved February 27, 2018.

Krueger, J. I. (2016). How not to believe. *Psychology Today Online*. https://www.psychologytoday.com/blog/one-among-many/201611/how-not-believe
Posted November 13, 2016, retrieved February 27, 2018.

Krueger, J. I. (2017). Reverse inference. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 110-124). New York, NY: Wiley.

Krueger, J. I., & Evans, A. M. (2013). Fiducia: Il dilemma sociale essenziale / Trust: The essential social dilemma. *In-Mind: Italy, 5*, 13-18. http://it.in-mind.org/article/fiducia-il-dilemma-sociale-essenziale

Krueger, J. I., Evans, A. M., & Heck, P. R. (2017). Let me help you help me: Trust between profit and prosociality. In P. A. M. Van Lange, B. Rockenbach, & T. Yamagishi (Eds.). *Social dilemmas: New perspectives on trust* (pp. 121-138). New York, NY: Oxford University Press.

Krueger, J. I., Freestone, D., & MacInnis, M. L. (2013). Comparisons in research and reasoning: Toward an integrative theory of social induction. *New Ideas in Psychology, 31,* 73-86. http://dx.doi.org/10.1016/j.newideapsych.2012.11.002

Krueger, J. I., & Massey, A. L. (2009). A rational reconstruction of misbehavior. *Social Cognition, 27*, 785-810. doi:10.1521/soco.2009.27.5.786

Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence to social norms. *Negotiation & Conflict Management Research, 1*, 31-52. doi:10.1111/j.1750-4716.2007.00003.x

Kruglanski, A. (2013). Only one? The default interventionist perspective as a unimodel – commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science, 8*, 242-247. https://doi.org/10.1177/1745691613483477

Kull, S., Ramsay, C., & Lewis, E. (2003). Misperceptions, the media, and the Iraq war. *Political Science Quarterly*, *118*(4), 569–598.

Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology, 36*(6), 635-642. http://dx.doi.org/10.1037/0022-3514.36.6.635

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444. doi:10.1038/nature14539

Lewandowskyy. S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing – therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science, 24*, 622-633. https://doi.org/10.1177/0956797612457686

Lewandowsky, S., Stritzke, W. G. K., Freund, A. M., Oberauer, K., & Krueger, J. I. (2013). Misinformation, disinformation and violent conflict: From Iraq and the "War on Terror" to future threats to peace. *American Psychologist, 68,* 487-501. doi: 10.1037/a0034515

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of*

Personality and Social Psychology, *37*, 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

McCright, A. M., & Dunlap, R. E. (2011). Cool dudes: The denial of climate change among conservative white males in the United States. *Global Environmental Change*, *21*, 1163–1172. https://doi.org/10.1016/j.gloenvcha.2011.06.003

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502–517. http://dx.doi.org/10.1037/0033- 295X.115.2.502

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in the Cognitive Sciences, 18*, 566-570. https://doi.org/10.1016/j.tics.2014.09.007

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision-making. *Cognition, 49*, 123-163. https://doi.org/10.1016/0010-0277(93)90038-W

Pennycook, G., Cannon, T. D., & Rand, D. G. (2017). Prior exposure increases perceived accuracy of fake news. *SSRN Electronic Journal*.

Pennycook, G., & Rand, D. G. (2017). Who falls for fake news? The roles of analytic thinking, motivated reasoning, political ideology, and bullshit receptivity. *SSRN Electronic Journal*. Retrieved from https://ssrn.com/abstract=3023545

Posner, M. I. (1973). *Cognition: An introduction*. Glenview, ILL: Scott, Foresman.

Rand, D. R. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science, 27*, 1192-1206. https://doi.org/10.1177/0956797616654455

Rasmussen C. E. (2004). Gaussian processes in machine learning. In O. Bousquet, U., von Luxburg, & G. Rätsch (eds.), *Advanced lectures on machine learning. Lecture notes in computer science* (Vol 3176) Heidelberg, Germany: Springer-Verlag.

Rieskamp, J., & Otto, P. E. (2006). SSL: A Theory of How People Learn to Select Strategies. *Journal of Experimental Psychology: General, 135*, 207-236.

Rothbart, M. (1981/2015). Memory processes and social beliefs. In D. L. Hamilton (ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145-182). New York, NY: Psychology Press.

Rousseau, J.-J. (1754/1984). *A discourse on inequality*. (M. Cranston, Trans.). New York, NY: Penguin.

Schank, R. C., & Abelson, R. P. (1995). Knowledge and memory: The real story. In R. S. Wyer, (ed.), *Advances in social cognition* (Vol. 8, pp. 1-85). Hillsdale, NJ: Erlbaum.

Shamas, V. (2018). *Deep creativity: Inside the creative mystery*. New York, NY: Morgan James.

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on Learning Representations*. http://arxiv.org/abs/1409.1556.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26. http://dx.doi.org/10.1111/1529-1006.001

Tetlock, P. E. (2003). Thinking the unthinkable: sacred values and taboo cognitions. *Trends in Cognitive Science, 7*, 320-324. https://doi.org/10.1016/S1364-6613(03)00135-9

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131. doi: 10.1126/science.185.4157.1124

von Horváth, Ö. (1937/2015). *Jugend ohne Gott*. Berlin, Germany: Holzinger.

Vyse, S. (1997/2014). *Believing in magic: The psychology of superstition*. New York, NY: Oxford University Press.