How gullible are we? A review of the evidence from psychology and social science

Hugo Mercier
CNRS

Abstract

A long tradition of scholarship, from ancient Greece to Marxism or some contemporary social psychology portrays humans as strongly gullible—wont to accept harmful messages by being unduly deferent. However, if humans are reasonably well adapted, they should not be strongly gullible: they should be vigilant towards communicated information. Evidence from experimental psychology reveals that humans are equipped with well-functioning mechanisms of epistemic vigilance. They check the plausibility of messages against their background beliefs, calibrate their trust as a function of the source's competence and benevolence, and critically evaluate arguments offered to them. Even if humans are equipped with well-functioning mechanisms of epistemic vigilance, an adaptive lag might render them gullible in the face of new challenges, from clever marketing to omnipresent propaganda. I review evidence from different cultural domains often taken as proof of strong gullibility: religion, demagoguery, propaganda, political campaigns, advertising, erroneous medical beliefs, and rumors. Converging evidence reveals that communication is much less influential than often believed—that religious proselytizing, propaganda, advertising, etc. are generally not very effective at changing people's minds. Beliefs that lead to costly behavior are even less likely to be accepted. Finally, it is also argued that most cases of acceptance of misguided communicated information do not stem from undue deference, but from a fit between the communicated information and the audience's preexisting beliefs.

Keywords: gullibility; epistemic vigilance; trust.

Are we too easily influenced by what people tell us? Do we accept messages that turn out to be harmful to us, even when we should have known better? Do we defer too easily to authority figures? In other words, are we gullible? Many scholars, from ancient philosophers to contemporary psychologists, have claimed that the answer is yes. I will argue that people are much less strongly gullible than is commonly believed. They are endowed with efficient mechanisms of epistemic vigilance that allow them to evaluate communicated information (Clément, 2006; Sperber et al., 2010).

The belief that most people are gullible has played a crucial role in Western intellectual thought. Ancient Greek philosophers' distrust of democracy was justified by their fear of demagogues, and the gullibility it implies on the part of the populace. As early as the fifth century BCE, Heraclitus wondered

> What use are the people's wits
> who let themselves be led
> by speechmakers, in crowds,
> without considering
> how many fools and thieves
> they are among, and how few
> choose the good? (Heraclitus, 2001 fragment 111)

The risk that a demagogue would be able to confound crowds and create a tyranny was one of Plato's main arguments against democracy (see Stanley, 2015; for Thucydides, see Ober, 1993). Even Aristotle warned that "democracy [can be] subverted by the wickedness of the demagogues" (*Politics*, Book 5, Chapter 5).

Since then, the fear of demagogues has maintained a strong grip on Western political thought. It played an important role in epochal political debates such as the framing of the U.S. Constitution (Signer, 2009; Tulis, 1987) and the reaction to the French Revolution (Femia, 2001). In the eighteenth century the risk that demagogues would mislead the masses was still "political philosophy's central reason for skepticism about democracy" (Stanley, 2015, p. 27).

Starting mainly in the eighteenth century, an opposite tradition developed. This tradition holds that it is not the people who revolt who are gullible, but those who fail to revolt. The Christian church was criticized for making people tolerate their plight: "It delivers mankind into [the] hands of [despots and tyrants] as a herd of slaves, of whom they may dispose at their pleasure" (Holbach, 1835, p. 119). This view was expanded and further articulated in the nineteenth century. In *The German Ideology*, Marx and Engels claimed that the dominant classes were able to impose their ideas to the whole of society: "the ideas of the ruling class are in every epoch the ruling ideas" (Marx & Engels, 1970, p. 64). Throughout the middle ages, peasants supposedly accepted the religious and feudal values of their oppressors (Poulantzas, 1978). Since the industrial revolution, workers have been thought to endorse the bourgeois values of the capitalists exploiting them (Marx & Engels, 1970). The 'dominant ideology thesis' has also been used to explain, for instance, the acceptation of the Indian caste system by people trapped in low caste or Dhalit status (Dumont, 1980). In each case, subordinate groups seem to be fairly gullible, as they are supposed to interiorize values that do not serve their interests.

Nowadays, attributions of gullibility are often used to explain political extremes: How propaganda provided the Nazi with obedient soldiers and the Communists with subservient workers, how American voters could have been taken in by Huey Long or Joseph McCarthy (Signer, 2009). Psychological experiments, most famously the Asch 'conformity experiments'

(Asch, 1956) and the Milgram 'obedience experiments' (Milgram, 1974), have been used to explain these events, in particular Nazi Germany. The common interpretation of these experiments, along with other results from psychology and the social sciences, has led to widespread claims that people are gullible, claims that can be found in academic journals (Gilbert, Krull, & Malone, 1990), in psychology textbooks (D. G. Myers, 2011; Zimbardo, Johnson, & McCann, 2006), in sociology textbooks (see Schweingruber & Wohlstein, 2005), in books by psychologists (Gilbert, 2006; Kahneman, 2011), philosophers (Brennan, 2012; Stanley, 2015), economists (Akerlof & Shiller, 2015; Chang, 2014), and other scholars (e.g. Dawkins, 2009; Herman & Chomsky, 2008).

From this very brief overview, it should be clear that the belief that people are gullible is not a straw man. It has been widely held by scholars throughout Western intellectual history, and it is still widespread. Yet I will argue that it is largely wrong.

The form of gullibility that is the main target here has the three following traits. First, it views gullibility as widespread: people would be very often fooled into accepting empirically unfounded messages. Second, it views gullibility as often applying to costly beliefs, beliefs that lead to painful rituals, expensive purchases, risky rebellions, or harmful complacence. Third, it views gullibility as being mostly source-based: stemming from the undue influence of focal sources, often authority figures, be they religious leaders, demagogues, TV anchors, celebrities, etc. Most accusations of gullibility reviewed above share these traits. I will refer to this view of gullibility as *strong gullibility*.

Evolutionarily, strong gullibility would be a puzzle, as revealed by the following simple argument. Communication can be evolutionarily stable only if it benefits both senders and receivers (Maynard Smith & Harper, 2003; Scott-Phillips, 2014). If senders do not benefit, they evolve to stop sending. If receivers do not benefit, they evolve to stop receiving. Gullibility entails frequent costs for receivers, as they accept false or misleading information. To the extent that human communication is adaptive (see, e.g. Pinker & Bloom, 1990; Scott-Phillips, 2014), humans should not, on average, be gullible.[1] On the contrary, they should skillfully discriminate harmful from beneficial information. The mechanisms that perform this function in humans have been dubbed mechanisms of *epistemic vigilance* (Sperber et al., 2010). Strong gullibility is clearly antithetical to well-functioning epistemic vigilance.

I will not argue that humans always make rational decision when it comes to evaluating communicated information. Instead I will try to make a case that people are equipped with a set of mechanisms of epistemic vigilance that are, on the whole, well adapted to the environment they evolved in (see, e.g. Tooby & Cosmides, 1992) or, in other words, that are ecologically rational (see, e.g. Gigerenzer, 2007). However, the environment mechanisms of epistemic vigilance would have adapted to is substantially different from the contemporary environment of most humans. Humans could thus be strongly gullible now in spite of the evolutionary argument sketched above.

In the present article, I review the evidence pertaining to human gullibility and vigilance. In the first half of the article, I briefly outline how the main mechanisms of epistemic vigilance should function, and review evidence from experimental psychology regarding their

---

[1] Note that this should be a weighed average. For instance, accepting many false but inconsequential messages could be balanced by the acceptation of a few messages that substantially increase our fitness. Still, the common view of gullibility depicted here often applies to the acceptation of messages that have significant consequences for those who accept them, from costly religious behaviors to engaging in risky rebellion or, on the contrary, accepting one's dire economic circumstances. The importance of this weighing as a function of the significance of the message is further elaborated in the introduction to the second part.

functioning (convergent evidence regarding other mechanisms of social influence can be found in Dezecache, 2015; and Dezecache, Mercier, & Scott-Phillips, 2013 for emotions; and in Morin, 2015 for imitation). The second half covers evidence from outside the lab, relying on different fields of social science.

Two literatures will not receive proper consideration. First, the philosophical literature on trust and social epistemology (e.g. Craig, 1999; Goldman, 1999; Williams, 2002). Even though this literature has played a role in shaping contemporary debates on the issues at hand, the current article aims at reviewing empirical rather than theoretical work. Second, epistemic vigilance can fail in two broad ways: through gullibility (accepting too much information), or conservatism (rejecting too much information), and I focus here on the former.

1 Psychology and gullibility

1.1 Plausibility checking

Given the relative rarity of lies (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996), and the benignity of many lies (DePaulo & Kashy, 1998), having an instinct to trust what others tell us seems to make sense (Reid, 2000). Cognitively, this might translate into mechanisms which treat trust as a default: these mechanisms would automatically accept communicated information, and only revise this initial acceptation later on, if necessary (Gilbert et al., 1990). The logic of this argument is flawed. What matters is not the current amount of abuse (in this case, lies and manipulation), but the potential for abuse. A system in which information is initially accepted and later potentially revised easily lends itself to abuse. In particular, if the revision stage were perturbed, the receiver would end up accepting information she might have otherwise rejected—including false or harmful information. Short of divine intervention—which was Reid's solution—it is not clear what would keep senders from evolving abilities to abuse such a design. This abuse would threaten the stability of communication.

An alternative design can be suggested which builds on how linguistic communication works. In the process of understanding linguistic communication receivers have to tap into their background beliefs (e.g. Kintsch, 1994; Sperber & Wilson, 1995). For instance, for John to understand Maria's utterance that "There is a green elephant in the yard," he has to activate beliefs about elephants and yards. If there are inconsistencies between John's background beliefs and what Maria intends to communicate, they must emerge at this stage. If they do not, then Maria's statement has simply not been understood, and could therefore not have its intended effect. Crucially, it has been argued that during this inferential process, communicated messages are encapsulated in a metarepresentational format, for instance of the form "Maria means that there is a green elephant in the yard" (Sperber, 1994; Sperber & Wilson, 1995). They would remain so encapsulated, and thus neither rejected outright not incorporated with our other beliefs, until they are found to be believable enough.

Inconsistencies between background beliefs and novel information easily lead to belief updating. If John *sees* a green elephant in the yard, he updates his beliefs accordingly. John can afford to do this because his perceptual and inferential mechanisms do not attempt to mislead him. By contrast, in the case of communicated information, the honesty of the sender is open to question. This means that communicated information that is inconsistent with a receiver's background beliefs should be, on average, less likely to lead to belief revision than similar information obtained through perception (in the absence of contrary evidence provided by trust or arguments, see below, and Sperber et al., 2010). We would thus rely on plausibility checking, a mechanism that detects inconsistencies between background beliefs

and communicated information, and that tends to reject communicated information when such inconsistencies emerge.

Plausibility checking mostly relies on mechanisms that are already necessary for understanding communication, dispensing with the need for a costly additional evaluation process. Moreover, the extent of the evaluation by plausibility checking is commensurate to the potential influence of the communicated information. The more inferences are drawn from a piece of communicated information, the more it can affect the receiver. But drawing more inferences requires activating more background beliefs, so that the information is also more thoroughly checked. Finally, plausibility checking would operate as the information is being understood (and maintained in a metarepresentational context), without the need to accept the information first, only to reject it later if necessary. It would thus not lend itself to the abuse that a default trust mechanism invites.

There is substantial evidence that people detect inconsistencies between their background beliefs and communicated information, that such inconsistencies tend to lead to reject communicated information, and that information that is more inconsistent with one's prior beliefs is more likely to be rejected. This evidence has been gathered in different literatures (with different experimental paradigms which will be described in the next section): persuasion and attitude change (Cacioppo & Petty, 1979), advice taking (Bonaccio & Dalal, 2006; Yaniv, 2004), information cascades (March, Krügel, & Ziegelmeyer, 2012; Weizsäcker, 2010) and development of trust in testimony (Clement, Koenig, & Harris, 2004; Koenig & Echols, 2003). For instance, in a typical advice taking experiment, participants have to form an opinion about a given question—typically a numerical opinion, such as 'how much does the individual in this picture weighs?' They are then confronted with the opinion of another participant, and they can revise their opinion on this basis. When no relevant factor, such as expertise, differentiate the participant receiving the advice from the participant giving the advice, the advice tends to be heavily discounted in favor of the participant's own opinion (e.g. Yaniv & Kleinberger, 2000). Moreover, the advice is more discounted when it is further away from the participant's own opinion (Yaniv, 2004).

These results, however, do not require a mechanism that checks plausibility as understanding occurs. It could still be that the communicated information is accepted, only to be rejected later, as it is more thoroughly checked. Indeed, several experiments suggest that this is precisely what happens (Gilbert et al., 1990; Gilbert, Tafarodi, & Malone, 1993). In these experiments, participants were shown statements followed by an indication of whether the statements were in fact true or false. For instance, participants could be told that in the Hopi language, "A monishna is a star," and then that this statement was false. In some trials, people's attention was disturbed as they were being told that the statement was true or false. On these trials, participants did not make random mistakes. Instead, they were more likely to think that the statements were true when they had been false, than false when they had been true. This suggests that, by default, participants treated the statements as true, and when the processing of the message telling that they were false was disturbed, the statements were simply treated as true—a truth bias.

However, there was nothing in the participants' background knowledge that could have led them to doubt any of the statements used in these experiments. Plausibility checking gives the participants no reason to reject the statement that, in the Hopi language, "A monishna is a star." Moreover, participants had no reason to mistrust the source of the statement (ultimately, the experimenter). The acceptance of the statement could therefore be deemed safe. When the statements are inconsistent with the participants' background beliefs, there is not truth bias. Instead, the statements are rejected from the start (Richter, Schroeder, & Wöhrmann, 2009).

The statements do not even have to be implausible to be rejected, they simply have to be relevant if false, in which case they tend to be treated as false by default (Hasson, Simmons, & Todorov, 2005; see also Street & Richardson, 2015a).

This evidence shows that there is no truth bias such that communicated information would be processed by default as being true. Instead, people engage in plausibility checking. Communicated information that is inconsistent with prior beliefs is detected and tends to be rejected. Since it makes it difficult for senders to change receivers' mind, an advantage of plausibility checking is its cautious conservatism. Note that this does not preclude effective communication of novel and useful information. For instance, the correct solution to a Eureka type problem is a novel and useful piece of information that would pass plausibility checking. Still, the conservatism of plausibility checking is also a drawback: in many cases, receivers would be better off accepting communicated information that challenges their prior beliefs. Other mechanisms of epistemic vigilance evolved in part to address this issue. These mechanisms provide receivers means to discriminate beneficial and harmful messages that can supersede plausibility checking.

1.2 Trust calibration

Trust calibration is one of the two mains mechanisms that can supersede plausibility checking (the other being argumentation, see below). Thanks to trust calibration, receivers can accept messages whose content is inconsistent with their prior beliefs. Trust calibration functions through two main mechanisms: cues to trustworthiness and commitment tracking. These mechanisms interact, but they obey different logics.

Receivers use a wide variety of cues to infer senders' trustworthiness. Some cues relate to the competence of the sender. A competent sender is a sender who is likely to have formed reliable beliefs. Cues to competence can be traits of senders such as dispositions (intelligence, diligence), or acquired expertise (being skilled in mechanics, say). Cues to competence can also be local, such as differences of perceptual access (who witnessed the crime).

Other cues relate to the sender's benevolence. A benevolent sender is a sender who is likely to send messages that positively take into account the receivers' interests (Barber, 1983). Thus, benevolence entails more than the absence of lying. If a sender sends a message that only benefits the sender, and not the receiver, she would not be deemed benevolent, even if the message is not an outright lie. For instance, a friend who recommends a restaurant on the basis of preferences she knows not to be shared by her audience would not be very benevolent. Like cues to competence, cues to benevolence can be traits, stable features of senders that make them more likely to be benevolent towards the relevant receiver (relatedness to the receiver, say). Cues to benevolence can also be local. In particular, attention should be paid to how the interests of the sender are served by the acceptation of the message being evaluated. Self-interested messages should arouse suspicion.

Coalitions have played an important role in human evolution (Tooby, Cosmides, & Price, 2006). Members of the same coalition tend to share more interests than random individuals. As a result, they should be more benevolent towards one another. Cues to coalition affiliation could thus be used as cues to trustworthiness. Some cues to coalition affiliation—such as physical appearance—seem relatively superficial, but they are easily trumped by more reliable cues, when such cues are available (Kurzban, Tooby, & Cosmides, 2001). We should expect the use of cues to coalition affiliation as cues to trustworthiness to follow the same pattern: use of more superficial cues when no other cues are available, use of the more reliable cues when they are available.

The other main mechanism on which trust rests is commitment. When a sender commits to a message, and the receiver finds the message to be unreliable, she imposes a cost on the sender, typically in the form of reduced trust (which entails a reduced ability for the sender to influence the receiver in the future). This potential cost provides an incentive for the sender to avoid sending unreliable messages. In turn, this incentive provides the receiver with a reason to believe the sender. Senders can rely on a variety of mechanisms to express their degree of commitment, from markers of confidence to explicit promises.

I now briefly review literature showing that receivers take cues to trustworthiness into account in an overall sensible way, and that they lower their trust in senders who were committed to messages that proved unreliable. I pay particular attention to studies suggesting that we put more weight than we should on the wrong cues. In each case, I argue that in fact the use of cues of trustworthiness and commitment is quite sound. The ideal presentation would consider the relevant cognitive mechanisms one by one, incorporating evidence from all the relevant fields (social psychology, developmental psychology, judgment and decision making) for each mechanism. However, given the current organization of the literature, and the fact that each field relies on different experimental paradigms, I have preferred to follow the standard disciplinary boundaries.

1.2.1 Persuasion and attitude change

The most relevant experiments carried out in this field follow this template: Participants attitude's towards a given object (a change in university policy, say) is measured before and after they read or hear a message. Two characteristics of the message are manipulated: how good its arguments are, and its sender (e.g. how expert the sender is in the relevant domain). The motivation of the participants to evaluate the conclusion is also frequently manipulated (for instance, the change in university policy affects the participants in one condition but not the other) (for review, see Petty & Wegener, 1998). This literature has shown that participants can take into account relevant sender cues such as expertise (e.g. Petty et al., 1981) or disinterestedness (Eagly, Wood, & Chaiken, 1978).

Participants have also been shown to take seemingly less sound cues into account, such as similarity (advantage to senders similar to the receivers Brock, 1965) or attractiveness (advantage to attractive senders, e.g. Chaiken, 1980). However, these cues typically only increase acceptance for messages that have little relevance to the receivers. When the messages are more relevant, positive cues of similarity and attractiveness (i.e. the sender is similar to the receiver or attractive) make receivers pay more attention to the message, which means that receivers only accept messages if they also contains strong arguments, and that if the messages contain weak arguments, they are more likely to damage the sender's credibility (for similarity, see Mackie, Worth, & Asuncion, 1990; for attractiveness, see Puckett, Petty, Cacioppo, & Fischer, 1983). Similar examples related to other sender characteristics can be found in the review of Petty and Wegener (1998).

1.2.2 Advice taking

In a typical advice taking study, participants are asked to make an estimate (typically a numerical estimate, such as "how much does this person weigh"), they are provided with the answer of another participant, and they can revise their estimate in light of this answer (for review, see Bonaccio & Dalal, 2006). The weight participants put on the other participant's opinion is influenced by relevant cues to trustworthiness, such as how expert a sender is relative to other senders and to the participant (e.g. Harvey & Fischer, 1997). Moreover,

participants quickly learn to discount more heavily the opinions of senders whose estimates have proven unreliable (Yaniv & Kleinberger, 2000).

1.2.3 Memory suggestibility

In a famous series of experiments, Loftus (1975) demonstrated that participants could be led to believe they remembered seeing something when in fact this memory had been suggested to them after the relevant event. For instance, in the context of a mock trial, participants had to watch a short movie of a car crash. After they had seen the movie, they were asked a series of questions which included, for some participants, the following: "Did you see the children getting on the school bus?" A week later, a quarter of those participants thought they remembered seeing a school bus in the movie, even though there had been no school bus. Participants who had not been asked this question were less likely to remember seeing a school bus.

Later experiments, however, revealed that for the presupposition (here, that there was a school bus) to have such an effect on participants' memory, it had to come from a source the participants had no reason to distrust—as in the original experiment. When the source was self-interested (such as the driver who had caused the incident), the presupposition was not integrated into the participants' memory (Dodd & Bradshaw, 1980; Echterhoff, Hirst, & Hussy, 2005). Other social influences on memory have been shown to be appropriately tailored to the source's competence. For instance, an experiment measured how much a discussion between two participants about some stimuli influenced the participants' memory of the stimuli. The influence each participant had on the other depended on the access they believed the other had had to the stimuli: if they believed the other had seen the stimuli longer than they had, they were more influenced, and vice versa (Gabbert, Memon, & Wright, 2007). On the whole, the evidence suggests that people's memory are not unduly influenced by others' testimonies (Jaeger, Lauris, Selmeczy, & Dobbins, 2012).

1.2.4 Development of trust in testimony

Most experiments investigating the development of trust in testimony follow one of these two templates. Either an informant gives the child some information that might contradict the child's prior beliefs, or several (usually two) informants provide contrary information on a topic about which the child has no prior belief. These experiments have revealed that young children (typically preschoolers) can take into account a wide variety of sensible cues to trustworthiness (for reviews, see Harris, 2012; Mills, 2013). Even infants (1- to 2-year-olds) can use some cues, such as expertise (for review, see Harris & Lane, 2013). While most of the cues used by preschoolers are sound (e.g.: past accuracy, Clément, Koenig, & Harris, 2004; benevolence, Mascaro & Sperber, 2009; perceptual access, Robinson, Champion, & Mitchell, 1999), others might appear to be less so (e.g. gender, Ma & Woolley, 2013). Regarding these apparently less sound cues, it should be noted that in the absence of other cues, they might make sense as cues to coalition affiliation. Moreover, they are usually trumped by sounder cues when sounder cues are present (for gender, Taylor, 2013; Terrier, Bernard, Mercier, & Clément, submitted).

1.2.5 Consensus and majority

In a series of experiments, Asch (1956) asked participants to complete a simple perceptual task with an unambiguously correct answer. However, before the participants could give their answer, they were preceded by several confederates who sometimes all gave the same wrong answer. The now standard way of describing these results is to emphasize that on a third of

the trials the participants ignored their perception and followed the consensus (see, Friend, Rafferty, & Bramel, 1990; Griggs, 2015). Asch, however, had intended the experiment as a demonstration of people's ability to ignore the pressure of the consensus (e.g. Asch, 1956, p. 10)—which is what participants did on two thirds of the trials. Twenty-five percent of participants never followed the consensus, while only 5% consistently followed it. Moreover, when participants gave their answers in writing, instead of voicing the answers out loud in front of the confederates, they only followed the consensus on 12.5% of the trials (Asch, 1956, p. 56). This suggests that on most of the trials in which participants had followed the consensus in the original experiment, they had not genuinely believed their answers to be correct. This latter result, suggesting that conformity is more normative than informative, has since been replicated, including conceptual replications with young children (Corriveau & Harris, 2010; Haun, Rekers, & Tomasello, 2014; Haun & Tomasello, 2011).

Majority effects tend to be even weaker than consensus effects. In Asch's experiments, when a single confederate providing the correct answer was present, participants followed the majority on only 5.5% of the trials (Asch, 1951). Many political science studies show weak or conflicting majority effects (for review, see Mutz, 1998). When people are told that the majority of the population holds a given opinion, this can affect their opinions, but it does not do so in a consistent manner: sometimes bandwagon effects (in which people shift towards the majority opinion) are observed, sometimes underdog effects (in which people shift towards the minority opinion) are observed (Mutz, 1998, p. 189ff). Moreover, bandwagon effects are typically observed only for issues to which the participants are weakly committed, and when no better source of information is present (Mutz, 1998, p. 194; see also, Giner-Sorolila & Chaiken, 1997). Research on small groups, and on persuasion and attitude change, reach similar conclusions regarding the propensity of group members to align with the majority or the minority view: although the opinion of the majority tends to carry a stronger weight, minority views are sometimes more influential (e.g. Baker & Petty, 1994; Moscovici, 1976; Trouche, Sander, & Mercier, 2014).

Finally, it should be stressed that, in the absence of sounder cues, majority is a robust cue to accuracy, so that it is far from unreasonable to follow it (Hastie & Kameda, 2005; Ladha, 1992). In line with this result, studies that have focused on informational conformity have found that it is taken into account, by adults and children, in a broadly rational manner: people tend to be more influenced by larger groups, by stronger majorities, and when they are less sure of their own opinions (Bernard, Harris, Terrier, & Clément, 2015; R. Bond, 2005; Campbell & Fairey, 1989; Gerard, Wilhelmy, & Conolley, 1968; McElreath et al., 2005; Morgan, Laland, & Harris, 2015; Morgan, Rendell, Ehn, Hoppitt, & Laland, 2012).

1.2.6 Deference to authority

Under the pretense of studying learning, Milgram (1974) had participants engage in what they thought was delivering a series of increasingly painful electric shocks to another participant (the learner). In fact, the learner was a confederate, and he was not shocked. Milgram found that 26 out of 40 participants (65%) complied with the experimenter's request and delivered the maximum shock of 450V. Interpreted as a demonstration that people blindly follow authority, these results have been vastly influential (e.g. Benjamin & Simpson, 2009).

However, the figure of 65% of participants blindly trusting the experimenter is misleading (even though uncritically reported in most psychology textbooks, Griggs & Whitehead, 2015). It was obtained in one of the 24 experiments conducted. Most of these conditions, including some that only involved minor variations such as a change of actors, yielded lower rates of compliance (conveniently reported in Perry, 2013, p. 304ff). More importantly, nearly

half of the participants reported not being sure that the learner was really receiving the shocks, which means that they might not have fully trusted the most crucial part of the setup (Milgram, 1974, p. 172). The participants who expressed no such doubt were less likely to comply. As a result, only 27% of participants can be said to have entirely trusted the experimenter—those who were sure the setup was genuine and who then complied. Even these participants were not blindly following authority. They deployed a wide variety of tactics to avoid complying, from various forms of verbal resistance (Hollander, 2015), to surreptitiously helping the learner, or trying to fake pressing the switch administering the shocks (Perry, 2013, p. 107 citing Milgram).

Crucially, compliant participants were not following just any authority. They were following the authority of science (Haslam & Reicher, 2012). Making the experiment look less scientific—by moving it away from Yale for instance—reduced compliance rates (Perry, 2013, p. 310, citing Milgram). Arguments that reminded participants of the scientific nature of the situation were generally necessary to produce compliance, while direct orders had close to no effect (Burger, Girgis, & Manning, 2011; Haslam, Reicher, & Birney, 2014).

That the deference demonstrated in Milgram's experiments is specifically deference to science—and thus a rational deference in many cases—fits with other results. Research carried out in the fields reviewed above show that adults (e.g. Harvey & Fischer, 1997; Petty, Cacioppo, & Goldman, 1981) and children (Landrum, Mills, & Johnston, 2013; Lane & Harris, 2015), on the whole, appropriately calibrate their deference as a function of the expertise of the source. By contrast, a few studies suggest a degree of 'leakage,' so that the opinion of an expert in an irrelevant field can be as influential as that of an expert in the relevant field (e.g. Ryckman, Rodda, & Sherman, 1972). Such leakage, however, only happens when participants have little knowledge of, and little interest in the topic at hand (Chaiken & Maheswaran, 1994).

1.2.7 Lie detection

In a typical lie detection experiment, participants are shown videos of people speaking, and the participants have to determine whether these people tell the truth or lie. The standard interpretation of the results from this field paints people as fairly gullible: participants (including experts) take the wrong cues into account (Global Deception Research Team, 2006), they perform barely above chance (C. F. Bond & DePaulo, 2006), and they have a bias to believe that people say the truth—a truth bias (C. F. Bond & DePaulo, 2006; Levine, 2014). More recent results, however, have shown that the picture of our lie detection skills should not be quite as bleak. On the contrary, participants might be making the best of the limited information available in lie detection experiments (Street, in press).

In the present framework, it is possible to make predictions about what should be reliable cues to deception. Behavioral cues to deception such as gaze avoidance or fidgeting should be wholly unreliable. These behaviors are not costly to avoid, and there is thus no reason we would have evolved to keep emitting such cues when lying. By contrast, the content of the message should provide more reliable cues. Assuming, trivially, that most lies are false, and that most people's beliefs are (approximately) true, it follows that a lie is more likely than a true statement to be incoherent either with the receiver's beliefs and with beliefs the receiver attributes to the sender. Such incoherence can therefore be taken as a cue to deception.

Some lies, however, can be communicated in such a way that they are incoherent neither with the receiver's beliefs nor with beliefs the receiver attributes to the sender. This is true, for instance, of lies about one's personal history delivered to a stranger. In such cases, I would

argue that if the lie is believed, it is generally because the sender has committed to it. When receivers rely on senders' commitment, they do not have to catch the lie right away, only to remember who said what, and to trust less senders whose statements are later found to have been unreliable. The next section will provide evidence that receivers rely on commitment to calibrate their trust.

A sender's degree of commitment can be used as a cue to deception. On the whole, more precise statements commit their senders more (Vullioud, Clément, Scott-Phillips, & Mercier, submitted). For instance, the reputation of a sender should suffer more if she is found to have been wrong when saying "I saw Linda kiss another guy passionately" than "I think Linda might be cheating on her husband." When a sender is expected to be able to provide precise statements, failing to do so might be seen as an unwillingness to commit, which would then be a potential cue to deception.

Studies of the cues that correlate with deception bear out these predictions. Behavioral cues such as gaze avoidance or fidgeting do not correlate with deception. By contrast, statements that are implausible or illogical tend to be deceptive, as well as ambiguous statements that contain few details (Hartwig & Bond, 2011). Why, then, do participants seem to take the former—unreliable—type of cue into account (Global Deception Research Team, 2006) while discounting the latter—reliable—cues (Strömwall, Granhag, & Hartwig, 2004)? One possibility is that participants have little introspective access to which cues they use. Supporting this explanation, a meta-analysis that examined the cues people actually take into to tell truths from lies revealed a near perfect correlation (r .93) between the cues to deception that are reliable and the cues people actually take into account (Hartwig & Bond, 2011, p. 655). In other words, people seem to use cues to deception in a near optimal manner.

The Hartwig and Bond (2011) study suggested two ways to reconcile the optimal way in which people use cues, and the low accuracy with which they detect lies in most experiments. First, most experiments are set up in such a way that the only cues that are present are unreliable. When strong diagnostic cues are present—such as who has an incentive to lie— participants achieve near perfect accuracy (C. F. Bond, Howard, Hutchison, & Masip, 2013). Second, most experiments ask participants to make explicit judgments. Several studies suggest that participants unconsciously discriminate against lies and liars: even when they are not explicitly labeled as such, lies tend to affect receivers less than truthful statements (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997), and liars are treated as being more suspect (Anderson, DePaulo, & Ansfield, 2002).

It thus seems that when reliable cues to deception are present, people do take them into account. In the absence of such cues, the best strategy is to fall back on the estimated base rate of true vs. deceptive statements in the relevant context. Since, in everyday interactions, people overwhelmingly tell the truth (DePaulo et al., 1996), it makes sense to have a bias towards believing others' statements to be true (provisionally at least, while keeping track of people's commitments, and in the absence of cues to deception of even falsehood, such as that provided by plausibility checking). When the base rate of true statement drops, either because it is experimentally manipulated (Street & Richardson, 2015b) or because the context changes (e.g. police interviewing a suspect Meissner & Kassin, 2002), then people believe more statements to be lies.

1.2.8 Commitment

By contrast with the previous domains, there is no literature dedicated to testing whether receivers track senders' commitments and punish committed senders whose messages turn out

to be unreliable. However, several strands of research suggest that receivers perform these tasks well enough. When participants are asked to track who is committed to what position in a dialogue, they perform very well (Rips, 1998). When senders make explicit commitments—such as promises—and that the commitments are then found to have been unreliable—they break their promises—they suffer large drops in trust (Schlenker, Helm, & Tedeschi, 1973; Schweitzer, Hershey, & Bradlow, 2006). Receivers are also sensitive to more fine-grained expressions of commitment. One such way to express commitment is through confidence—e.g. saying "I'm really sure." Messages expressed with greater confidence tend to be taken into account more than messages expressed less confidently (e.g. Price & Stone, 2004). However, when the messages are found to have been unreliable, the reputation of the individual who sent the more confident message tends to suffer more (Tenney, MacCoun, Spellman, & Hastie, 2007; Tenney, Spellman, & MacCoun, 2008; Vullioud et al., submitted).

1.3 Reasoning

Reasoning can be defined here as the ability to find and evaluate reasons that can affect belief in a conclusion (see Mercier & Sperber, 2011). Reasoning can be private—in the case of ratiocination—or public—in the case of argumentation. A long philosophical tradition has warned against the dangers of fallacies and sophistry, claiming that people are too easily swayed by poor arguments, which might make them gullible. Experimental evidence suggests that these concerns are unfounded.

Several fields have studied how people evaluate others' arguments, using different means to manipulate the strength of arguments and to measure their effects. In the persuasion and attitude change literature mentioned above, researchers have mostly relied on informal criteria of argument strength. For example, a weak argument could be hearsay from an unreliable source while a strong argument could be relevant evidence from a reliable source. When the conclusion of the argument is of little import to the participants, arguments have close to no effect either way. When the conclusion of the argument is relevant to the participants, weak arguments are forcefully rejected (e.g., Petty & Cacioppo, 1979; for review, see Petty & Wegener, 1998). Either way, weak arguments do not have the intended effect on receivers (for similar evidence with young children, see Castelain, Bernard, Van der Henst, & Mercier, 2016; Koenig, 2012; Mercier, Bernard, & Clément, 2014).

Other researchers have relied on standard argument categories—*ad populum*, *ad verecundiam*, etc. (for review, see Hornikx & Hahn, 2012). They found that when flaws were introduced in arguments, people appropriately rated them as being weaker. For instance, participants find less compelling an argument from authority when the authority is not an authority in the relevant area, or has vested interests (Hoeken, Šorm, & Schellens, 2014). Similarly, continuous variations in some argument features can be predicted, using Bayes' rule, to increase or decrease argument quality. Participants react appropriately to these variations for several argument types (Hahn & Oaksford, 2007). For example, a Bayesian analysis predicts that among the following two arguments from ignorance, the former should be stronger than the latter (Hahn & Oaksford, 2007, p. 708):

Drug A is not toxic because no toxic effects were observed in 50 tests.
Drug A is not toxic because no toxic effects were observed in 1 test.

Participants rated the former argument as being stronger than the latter. This example demonstrates that some arguments that would be typically classified as fallacies (arguments from ignorance) are perfectly sensible, and that they are judged as such by participants. By contrast, when participants are provided with textbook fallacies (in this case, "Ghosts exist,

13

because nobody has proved that they do not"), they are much more critical (Hahn & Oaksford, 2007).

The studies mentioned above also show that the rejection of weak arguments does not stem from a blanket rejection of all arguments that challenge one's prior beliefs. Strong arguments are rated positively, and they influence participants, even when they run against prior preferences or beliefs (e.g. Petty & Cacioppo, 1979; Trouche et al., 2014; Trouche, Shao, & Mercier, submitted).

2 Culture and gullibility

The experimental psychology results reviewed above demonstrate that people are endowed with mechanisms of epistemic vigilance that work, in the laboratory at least, reasonably well. They evaluate messages based on their content, on various attributes of their source, and on the arguments provided to support them: they are, broadly, ecologically rational.

Most accusations of gullibility, however, have been leveled not based on data from experimental psychology, but on observations from outside the laboratory: demagogues using fallacious arguments, charlatans peddling snake oil, etc. It is possible that mechanisms of epistemic vigilance that function well in the laboratory fail otherwise, when they are confronted to years of indoctrination, marketing departments, or professional spin doctors. In the second part of this article I briefly review the various domains that have historically given rise to accusations of gullibility. This review suggests that the common view of strong gullibility is largely inaccurate.

I will also argue that most instances of gullibility are the outcome of content-based, rather than source-based, processes, and that they only indirectly bear on the working of epistemic vigilance. Mechanisms influencing the amount of epistemic trust to allot different sources are essential components of epistemic vigilance. When people are overly deferent towards some sources, epistemic vigilance is failing. Epistemic vigilance is also geared towards the content of communicated information, through plausibility checking and reasoning. However, these mechanisms rely on intuitions and beliefs that are not within the purview of epistemic vigilance *per se*. For instance, if a medieval peasant accepts the communicated belief that the Earth is flat, the reason this belief is accepted does not lay with faulty plausibility checking, since this belief is consistent with the peasant's observations. Instead, it lies with the observations, and the inferences that are intuitively drawn from them. So, even though the peasant might be said to be, in some sense, gullible, the mechanisms of epistemic vigilance— here, plausibility checking—are not directly to blame.

This does not mean that plausibility checking or reasoning cannot fail in their own right. People could be tricked by sophistry and various other types of manipulative messages to accept messages that are in fact largely inconsistent with their beliefs (see, e.g., Maillat & Oswald, 2009; Stanley, 2015). Still, in the review that follows, I will focus on cases in which people accept messages because they fit with their pre-existing beliefs or intuitions, so that the flaw rests with these beliefs and intuitions rather than with epistemic vigilance *per se*. One might say that all that matters is that mistaken beliefs are accepted, irrespective of the specific cognitive mechanisms involved but, in the conclusion, I will argue that the distinctions drawn here between source-based and content-based processes is not only theoretically, but also practically significant.

Against strong gullibility, I presently review evidence from different literatures on the diffusion of communicated information to support the three following propositions. 1)

Communicated information is less influential than suggested by strong gullibility. 2) Communicated information is less harmful than suggested by strong gullibility. 3) Content-based processes play a more important role in the acceptance of empirically unfounded messages, and source-based processes a small role than suggested by strong gullibility.

The following review broaches many topics. This range comes at a cost in terms of the depth with which each issue is examined. As a result, I have tried to mostly incorporate results that are, as far as I can tell, commonly accepted within their respective field. The relevance of this review does not stem from developing novel interpretations, but from demonstrating the convergence between evidence accumulated in fields that have few contacts with each other, and even fewer opportunities for integration. When the interpretations put forward are disputed within the relevant community, I try to make this clear.

2.1 Religion

In particular since the enlightenment, people holding religious beliefs, or engaging in religious behaviors, have been prime targets for accusations of strong gullibility. Here I will argue that these accusations are largely misguided. To do so, I will rely on work in the cognitive anthropology of religion that has attempted to explain what makes some religious beliefs culturally successful. One of the first observations made in this literature is that religion is not a natural kind. Small human groups typically do not have a category of beliefs and practices that corresponds to what most people identify as religion now (Boyer, 2001). It is to their beliefs and practices that I first turn.

2.1.1 'Religious' beliefs in small human groups

Human in small groups—groups of hunter-gatherers, chiefdoms—hold 'religious' beliefs such as beliefs in supernatural agents—local deities or dead ancestors for instance (Boyer, 2001). However, a common anthropological observation is that members of these communities appear to be only weakly committed to these beliefs (e.g. Bloch, 1989; Evans-Pritchard, 1971; for a similar observations about supernatural beliefs in Western societies, see Luhrmann, 1991). In particular, most of these 'religious' beliefs do not seem to be held in the same way as most other beliefs (Baumard & Boyer, 2013b; Sperber, 1975, 1997). In the vocabulary introduced by Sperber, they are held reflectively, not intuitively.

Most beliefs are held intuitively, meaning that they can freely interact with other beliefs and inferential mechanisms. By contrast, reflective beliefs remain embedded in a propositional attitude. Instead of holding the belief "there is a talking snake" in the same way as one would hold the intuitive belief "there is a snake," one would instead hold a belief such as "we believe there is a talking snake." The embeddedness of "there is a talking snake" stops it from interacting freely with other beliefs or inferential mechanisms (Sperber, 1997). For instance, people can believe there is a talking snake, without believing that real snakes are likely to talk, and without talking to them in the hope that they will reply (Barrett, 1997; Slone, 2007). This means that the cognitive costs (generation of inaccurate beliefs) and behavioral costs (maladaptive behaviors) stemming from reflective beliefs are much inferior to what they would be if the same beliefs were held intuitively. That these 'religious' beliefs are only held reflectively is consistent with the observation that the behaviors associated with these beliefs tend to not be very costly. For instance, elders might require animal sacrifices to mollify the spirit of a dead ancestor—but the sacrificed animal is eaten, not left to rot (see Boyer & Baumard, in press) (note that there are significant exceptions to this pattern, see conclusion).

Moreover, some cognitive anthropologists have suggested that the acceptation of these 'religious' beliefs is largely driven by content-based processes rather than source-based processes. More specifically, the cultural success of these beliefs would largely rest in the way they tap into a range of cognitive mechanisms (Sperber, 1975, 1996). For instance, successful beliefs in supernatural agents would spread in part because they are minimally counter-intuitive (Barrett, 2000; Boyer, 2001). A belief in a talking snake is counter-intuitive in that snakes do not talk. It is nonetheless intuitive in most ways: the behavior of the talking snake can be understood intuitively using folk biology and folk psychology. The talking snake needs to eat, engages in behaviors based on his beliefs and desires, etc. Such minimally counter-intuitive beliefs are interesting, memorable, and easy enough to understand, making them more likely to be culturally successful.

That such cognitive factors play a more important role than the authority of the relevant specialists—shamans or elders—in explaining the success of some 'religious' beliefs is suggested by two pieces of evidence. The first is the existence of strong cross-cultural patterns in the types of 'religious' beliefs found in small human groups (Boyer, 2001). If people were mostly following the idiosyncratic beliefs of the local shaman, one would expect much more variation. Second, the authority of the relevant specialists, such as shamans, typically does not extend beyond their domain of expertise, and they are not believed because they occupy a dominant place in the group (Boyer & Baumard, in press). Although these explanations are not consensual, it is thus plausible that the 'religious' beliefs that spontaneously emerge in small human groups are not good examples of strong gullibility. They may be widespread, but they tend to not be costly, and their success might be largely owed to content-based processes.

2.1.2 World religions

By contrast with deities of small human groups, the gods of contemporary, world religions (such as Christianity) are moralizing gods (Baumard & Boyer, 2013a; Bellah, 2011). These gods are thought to favor many beliefs and behaviors that do not seem to be in the short term interest of believers: fasting and sexual restraint, patience and delay of gratification, tithing and pilgrimages, etc. (Baumard & Boyer, 2013a; Baumard & Chevallier, in press; Bellah, 2011; Wright, 2009). Moreover, world religions typically have an established clergy proselytizing an established doctrine. It might thus seem that believers in moralizing gods have been tricked by clergies into accepting a variety of beliefs and behaviors that are costly for them but beneficial to the clergy or other dominant classes—a perfect example of strong gullibility.

However, some well-supported, even if controversial, explanations for the rise of world religions do not entail strong gullibility. In particular, Baumard and his colleagues (Baumard & Boyer, 2013a; Baumard & Chevallier, in press; Baumard, Hyafil, Morris, & Boyer, 2015) have argued that the beliefs characterizing world religions emerge instead because they serve the interests of their bearers. They have shown that the emergence of world religions during the Axial age (eighth to third century BCE) is best explained by rising affluence. As more people enjoyed a lifestyle above subsistence levels, and could expect to keep enjoying this lifestyle in the future, they shifted towards strategies that favor long-term investments. They then adopted a suite of beliefs that allowed them to justify these strategies towards those with less affluent lifestyles (Baumard & Chevallier, in press). These beliefs included some forms of sexual restrain and delay of gratification for instance.

Even if one accepts this account, does not the success of the early, affluent adopters to spread world religions suggest a large measure of gullibility on the part of the less affluent? Not

necessarily. First, rising affluence could have contributed to the spread of these religions by making them more congenial to an increasing number of people. Second, affluent people tend to yield more power, and, throughout history, they have used this power to preach their religious beliefs. They have done so both through positive incentives (e.g. charity) and negative incentives (penalties for paganism, heresy, apostasy, etc.). These incentives might have led a number of the less affluent to nominally accept the beliefs of world religions.

This account predicts that when affluence is not sufficient to make beliefs associated with a world religion congenial, then the acceptation of these beliefs should be superficial, selective, and resisted when costly. The official doctrine of world religions should either be rejected or transformed to be closer to the 'religious' beliefs that spontaneously emerge in small human groups, beliefs that the doctrine aims to replace. As an example, I now present some evidence of the failures of an official world religion doctrine: Christianity in the European Middle Ages (for more evidence, see, Abercrombie, Hill, & Turner, 1980; Delumeau, 1977; Le Bras, 1955; Stark, 1999; Thomas, 1971).

It is likely true that most people, at most times of the European Middle Ages, would have professed some Christian beliefs, and would have engaged in some behaviors sanctioned by the Church (Christenings, marriages, some mass attendance). However, the beliefs were often held superficially. As a thirteenth century Dominican preacher complained, people go to mass but they "carry nothing away with them but words" (Murray, 1974, p. 303). Moreover, professions of belief cannot be equated with acceptation of the doctrine, since these professions were often used strategically in heresies and other revolts, including in violent revolts against the clergy (Cohn, 1970).

Christian beliefs did not displace the beliefs they aimed at replacing. In particular, the 'religious' beliefs that spontaneously emerge in small human groups were still present throughout the Middle Ages (and at least the Early Modern period). People kept believing in supernatural agents, spells, and witches (Thomas, 1971). Not only did the Christian doctrine fail to replace such beliefs, but they heavily influenced Christian doctrine. The introduction of Saints is a good example: Saints are supernatural agents with whom one can interact for material benefits, much like local deities or the ancestors in ancestor cults (Boyer & Baumard, in press).

Turning to behavior, many nominally Christian behaviors can be explained by mere convenience. For instance, in a largely illiterate society, it was useful to have the clergy keep track of births and weddings. Other behaviors, such as pilgrimages or crusades seem costly but might have been motivated by material incentives. Pilgrimages could be the opportunity for "more sin, sometimes, than a participant committed in all the rest of the year put together" (Murray, 1974, p. 318). Poor people's crusades chiefly recruited individuals who had little to lose (no land, no work, little kin support) (Cohn, 1970; Dickson, 2007), and who stood to gain materially (these crusades relied heavily on looting) (Cohn, 1970; Dickson, 2007). Far from being the result of propaganda from the official Catholic clergy, people's crusades were regularly reproved by the Popes (Cohn, 1970; Dickson, 2007). Behaviors that were costly but bore no benefit in return were actively resisted. If the costs were high, resistance could be violent—as in the heretical movements that refused to pay the tithe (Ekelund, Hébert, Tollison, Anderson, & Davidson, 1997). Even if the costs were low—confession, penance, fasting—priests found it difficult to get people to comply (Abercrombie et al., 1980; Murray, 1974).

To sum up, it can be argued that most of the beliefs that characterize world religions have not been foisted on the population. They would instead serve well the interests of some

individuals—originally, some affluent members of society. Whether or not this is true, the historical evidence clearly reveals that many members of society nominally dominated by world religions only accept these religious beliefs partially, endorsing innocuous ones—often in the superficial form of "theological correctness" (Barrett, 1999; Slone, 2007), transforming others, rejecting those that would lead to costly behaviors. Thus, the spread of these beliefs does not constitute proof of strong gullibility.

2.2 Demagogues and charisma

The success of demagogues has sometimes been attributed to their charisma: "the governed submit because of their belief in the extraordinary qualities of the specific person" (Weber, 1958, p. 295; see also, e.g. Teiwes, 1984). Unless these extraordinary qualities were to correlate very well with the ability and willingness to make decisions that benefit the governed, submitting to the rule of charismatic leaders would be a form of strong gullibility. Even if it were true that demagogues propound many beliefs that end up being costly for the population, it can be argued that their success largely rests on content-based, rather than source-based processes. Under this explanation, demagogues are not successful because they are charismatic. They are deemed charismatic when they are successful, that is, when they give voice to the pre-existing beliefs and desires of the population (Kershaw, 1987; Worsley, 1957).

This explanation is supported by at least three arguments. The first is that many demagogues were not described as charismatic outside of their coterie before they rose to power (e.g. Kershaw, 1991). The second is that in order to rise to power, they have to adjust their rhetoric to their audiences (for instance, Hitler was forced to put much more weight on anti-communism than on anti-Semitism, Kershaw, 1987). The third is that even when they are in power, demagogues often fail to influence the population in the way they desire (see, Signer, 2009; Wang, 1995, this point is elaborated in the next section). It is thus plausible that demagogues generally do not constitute good examples of strong gullibility, since they do not owe their success chiefly to their personal traits, but rather to the messages they propound.

2.3 Propaganda

Even if demagogues cannot influence the masses through sheer charisma, they might be able to achieve this influence once they gain access to a modern state's propaganda apparatus. The efficacy of propaganda is suggested by professions and shows of support, from Nazi rallies to scenes of mourning for Kim Jong Il's funeral, which would constitute examples of strong gullibility.

These professions and shows of support, however, give a misleading impression of the success of propaganda. Although the content of propaganda has received much more attention than its efficacy, on the whole, the evidence available suggests that the efficacy of propaganda is generally too limited to constitute strong gullibility.

First, propaganda often fails. To take the example of Nazi propaganda, it failed to generate support for euthanasia of the handicapped (Kershaw, 1983a; Kuller, 2015), it largely failed to turn people into rabid anti-Semites (Kershaw, 1983b; Voigtländer & Voth, 2015), it failed to generate much liking for the Nazi party (Kershaw, 1983b, 1987), and it soon failed to make Germans more optimistic about the outcome of the war (Kallis, 2008; Kershaw, 1983a; for similar examples regarding Stalinist propaganda, see Brandenberger, 2012; Davies, 1997; Maoist propaganda, see Wang, 1995; North Korean propaganda, see B. R. Myers, 2011).

Moreover, many propaganda failures are avoided only because the propagandists adapt their messages to the actual or anticipated reaction of the population. This happened when the Nazis had to shift from optimistic to increasingly grim predictions about the outcome of the war (Kallis, 2008), when the Communist party under Stalin had to shift from internationalism to patriotism, and from abstract, impersonal historical forces to the cult of heroes (Brandenberger, 2012), when North Korean propaganda had to shift from boasting about higher standards of living to gloating about military strength (B. R. Myers, 2011). When propagandists fail to adapt to the population's beliefs, their efforts backfire, so that the main effect of propaganda is to decrease trust in propaganda and the government it emanates from (Chen & Shi, 2001; Davies, 1997; Kershaw, 1983a).

These failures suggest that the success of propaganda does not stem from its methods (domination of the media, repetition, fallacious arguments, etc.), but largely from the fit between its content and the population's preexisting beliefs. Propaganda's effectiveness is "heavily dependent on its ability to build on existing consensus, to confirm existing values, to bolster existing prejudices" (Kershaw, 1983a, p. 200). For instance, the effectiveness of Nazi anti-Semitic propaganda correlates with the intensity of prior anti-Semitic beliefs, but not with the intensity of the propaganda (Adena, Enikolopov, Petrova, Santarosa, & Zhuravskaya, 2015; Voigtländer & Voth, 2015). This seems to be largely true for other propaganda efforts as well (e.g. Wang, 1995).

The professions and shows of support for the regime, which are suggestive of propaganda's efficacy, should not be taken at face value, since they can be explained at least in part by incentives. Support for the regime is rewarded, opposition is harshly punished. Opposition could consist merely in failing to perform the Nazi salute (Kershaw, 1987), or pointing out some minor inconsistency in North Korean propaganda (Demick, 2010). Moreover, propaganda symbols are sometimes used by the population not with their intended sense but instead in a way that subverts their official meaning (Davies, 1997; O'Brien & Li, 2008; Scott, 1990). Finally, even genuine professions and shows of support have to be put in balance with professions and shows of disapproval, from failures to show up at rallies (Hoffmann, 1991, p. 120), absenteeism (Salter, 1983), and other tactics of passive resistance (Filtzer, 1986), to outright strikes and riots (Friedman, 2014).

To summarize, it is well attested that even massive propaganda efforts often fail. Shows of support for regimes that rely heavily on propaganda can in part be explained by the reliance on incentives, and they have to be contrasted with manifestations of disagreement which are often less visible. Moreover, it can be argued that when propaganda efforts do succeed it is because propagandists have tailored their messages to the beliefs and preferences of the population. The efficacy of propaganda could thus be largely content-based rather than source-based. It thus seems that few propaganda efforts could be characterized as examples of strong gullibility.

2.4 Mass media and political campaigns

In modern democracies, voters are sometimes thought to vote against their interests as a result of manipulations by the media and political campaigns, making for another potential instance of strong gullibility (Frank, 2007; Herman & Chomsky, 2008). In the early years of the twentieth century, a common view (Lasswell, 1972; Lippmann, 1965) was that mass media could easily sway public opinion. Far from supporting this view of a strongly gullible public, the first empirical studies revealed instead that the media and political campaigns only had "minimal effects" on the public's thoughts and in particular on its voting behavior (Klapper, 1960). Since, experimental studies have led to a reemergence of a belief in media effects,

albeit more subtle ones. Instead of directly telling people what to think, the media would tell people what to think about (agenda setting, Iyengar & Kinder, 1987), how to best understand issues (framing, Gamson, 1992), and what criteria to use in evaluating politicians (priming, Iyengar & Kinder, 1987).

Some experimental studies, however, likely overplay the role of the media outside the laboratory. In contrast with most experimental settings, people outside the laboratory do not passively receive political information, they actively seek it out—or, in most cases, do *not* seek it out. As a result, many people are simply not exposed to much political information that might influence them (Arceneaux & Johnson, 2013). More to the point, it also seems that the people who are exposed to such information are those least likely to be influenced (Arceneaux & Johnson, 2013). Moreover, outside the laboratory people are often confronted with different sources of information—either from the media or through their personal networks. The different frames they are then exposed can cancel each other out (Chong & Druckman, 2007; Druckman, 2004).

Still, some effects seem to be robust even outside the laboratory. In particular, people often adopt the opinions of their favored candidate or party ("party cues," see, Lenz, 2013). Indeed, what was thought to be priming (i.e. voters supporting a candidate because of her position on a given issue made salient) is often better explained by the tendency of voters to adopt a favored politician's positions (Lenz, 2009). Party cues could be especially problematic as they extend to factual beliefs, so that for instance in the U.S. Democrats would tend to overestimate the performance of Democratic presidents, while Republicans would underestimate it. Such results, however, should not be taken at face value. Answers to surveys do not directly reflect what people think is true. Instead, they are in large part a way of displaying one's political leanings, so that it is unsurprising that the results are influenced by party cues (Bullock, Gerber, Hill, & Huber, 2013; Prior & Lupia, 2008). Even when the influence seems to be genuine, it mostly happens when people do not have more accurate information to ground their opinions on (Bullock, 2011; Druckman, Peterson, & Slothuus, 2013). Thus, party cues mostly affect non-attitudes, beliefs people hold only very superficially and that do not influence other opinions (Kuklinski, Luskin, & Bolland, 1991) or behaviors such as voting (Lenz, 2013). By contrast with these non-attitudes, when the public holds more steadfast opinions, it is politicians who tend to adapt their policies to the public's opinions (Lenz, 2013; Stimson, 2004).

Finally, to the extent that political campaigns and the media influence the outcome of elections, it is mostly by relaying accurate information. Election outcomes are largely based on the state of the economy and other significant variables (serious breaches of trust from politicians, wars, etc.) (Gelman & King, 1993; Kinder, 1998; Lenz, 2013). This suggests that whether the information comes from the media, other elites (Lupia, 1994), or personal networks (Katz & Lazarsfeld, 1955; Lazarsfeld, Berelson, & Gaudet, 1948), people calibrate their trust appropriately on the whole. The positive role of the media is also supported by the two following observation. First, people who trust the media less tend to have less accurate knowledge about economic and political facts (Ladd, 2011). Second, politicians facing constituents with more news source, and who are thus better informed about their actions, have to make more efforts to satisfy the constituents' interests (Besley, Burgess, & others, 2002; Snyder & Strömberg, 2010; Strömberg, 2004).

To sum up, voters are rarely influenced by the media and political campaigns in a way that is both significant (i.e. substantial change in opinion) and personally costly (for a review with some potential exceptions, see, DellaVigna & Gentzkow, 2010). In most cases, the influence of the media and political campaigns is small. When the influence is large, it generally leads

to more accurate knowledge of important facts, or to less accurate beliefs about matters of little significance. On the whole, research in political science suggests that the influence of political campaigns and of the media is positive, making for a better-informed public (for some potential exceptions, see, DellaVigna & Kaplan, 2007). Even in the cases in which the influence is negative, it is likely driven by content-based rather than source-based process: slant in media coverage is more strongly determined by demand—what people want to hear—than by supply—what the media would like the public to believe (Gentzkow & Shapiro, 2010) (unfortunately, this might lead to the spread of misleading beliefs, if people start out with mistaken preconceptions, as discussed in conclusion).

This brief survey relies heavily on data from the U.S., for which the most information is available. To the best of my knowledge, there is no indication that media effects are substantially different in other democratic countries. On the whole, the political science literature provides very little support for strong gullibility.

2.5 Advertising

Are customers influenced by ads into buying products they will not enjoy? Are they gullibly influenced by the sheer repetition of ads or the endorsement by irrelevant celebrities? Advertising could constitute another example of strong gullibility. However, the evidence quite clearly shows that this is not the case. Even though it is difficult to measure the effects of advertising, thanks to big data and meta-analyses, it is possible to get an estimate of advertising's average effects. Compared to tangible factors such as satisfaction with a product (Oliver, 2014) or price (Tellis, 2003), the effects of advertising are very small (Blake, Nosko, & Tadelis, 2014; DellaVigna & Gentzkow, 2010; Sethuraman, Tellis, & Briesch, 2011; Tellis, 2003; for political ads, see, Broockman & Green, 2014; Gerber, Gimpel, Green, & Shaw, 2011).

Even when consumers are affected by advertising, this does not make them automatically strongly gullible (or even gullible at all). On the whole, people seem to pay attention to sensible cues in advertising. The content of the ads matter more than sheer repetition (Van Den Putte, 2009); when celebrity endorsements work, it is usually when the celebrity is a trustworthy expert (Amos, Holmes, & Strutton, 2008; Spry, Pappu, & Bettina Cornwell, 2011); gratuitous sex and violence in ads decrease their impact (Lull & Bushman, 2015). More importantly, ads which exert a significant influence can either make the customer better off (if she discovers a new product she enjoys) or at least not worse off (if she switches between two quasi-identical products such as different brands of sodas, of cigarettes, etc.). There is no evidence of widespread gullibility in response to advertising: "the truth, as many advertisers will quickly admit, is that persuasion is very tough. It is even more difficult to persuade consumers to adopt a new opinion, attitude, or behavior" (Tellis, 2003, p. 32).

2.6 Erroneous medical beliefs

For most of history, people have held inaccurate beliefs about topics now covered by science—beliefs about the origin of the world, about astronomy, etc. Here I want to focus on a subset of these beliefs—medical beliefs—that is particularly relevant for two reasons: they are of great practical import, and it is often believed that they owe their success to the prestige of the physicians, shamans, or other specialists defending them. They could thus constitute a good example of strong gullibility.

Bloodletting is a salient example. One of the most common therapies for significant portions of Western history, inefficient at best and lethal at worse, it seems to have owed its success to

the authority granted the writings of Galen and other prestigious physicians (Arika, 2007; Wootton, 2006). This thus seems to be a blatant example of misplaced prestige bias (Henrich & Gil-White, 2001). However, ethnographic data reveal that bloodletting is a common practice worldwide, occurring in many unrelated cultures, on all continents, including in many cultures which had not been in contact with Westerners (Miton, Claidière, & Mercier, 2015). These ethnographic data, as well as some experimental evidence, suggests that bloodletting owes its cultural success to its intuitiveness: in the absence of relevant medical knowledge, people find bloodletting to be an intuitive cure (Miton et al., 2015). If this explanation is correct, trust would flow in the other direction: instead of bloodletting being practiced because it is defended by prestigious physicians, it is because some physicians practiced and defended bloodletting that they became prestigious. This explanation could easily be extended to the most common forms of therapies in pre-modern cultures, which all aim at removing some supposedly bad element from the body (laxatives, emetics, sudation, see, Coury, 1967)

Similarly, people would not refuse to vaccinate their children because they follow Jenny McCarthy or other prominent anti-vaxxers. Instead these figures would become popular because they attack a very counter-intuitive therapy (Miton & Mercier, 2015). This phenomenon would thus be similar to that of political or religious leaders who are mostly deemed charismatic and prestigious because they endorse popular positions. In neither case would people be gullibly following prestigious leaders, instead they would simply be heeding messages they find appealing, and then conferring some prestige on those who defend them. The spread of misguided beliefs would thus mostly rest on content-based rather than source-based processes.

2.7 Rumors

Rumors are generally perceived negatively. People claim rumors are not credible (DiFonzo & Bordia, 2007), and the most prominent examples are of false rumors (that Barack Obama is a Muslim for instance). This impression, however, is misleading. In some contexts rumors are overwhelmingly true. Studies show that rumors circulating in the workplace are accurate at least 80% of the time (DiFonzo & Bordia, 2007, p. 146 and references within). In other contexts, such as that of catastrophic events (war, natural disaster) rumor accuracy can be much lower (ibid.) (although electronic media might help make even such rumors more accurate, see Gayo-Avello, Peter Gloor, Castillo, Mendoza, & Poblete, 2013).

To some extent, the same factors can explain both the accuracy and lack of accuracy of rumors. The two main factors that seem to be at play are rumor plausibility and the potential costs of ignoring the rumor. When rumors are transformed in the process of transmission, they converge towards a narrative that is deemed plausible by the members of the relevant population. This was well demonstrated in classic chain experiments in which stereotypes heavily biased rumor transmission (Allport & Postman, 1947). In this case, plausibility played against accuracy: the rumors became increasingly plausible for the participants, but because the original rumor violated stereotypes, increased plausibility translated into increased distortions. However, in many cases evaluations of plausibility play a positive role. In particular, it stops rumors deemed implausible (and which should, generally, be less likely to be true) from spreading. For instance, during World War II, negative rumors about Japanese Americans mostly spread in places where there were few Japanese Americans. When people had enough first hand knowledge of Japanese Americans they did not believe the rumors—even though they would have been more at risk if the rumors had been true (Shibutani, 1966). Even the spread of rumors surrounding such intense events as ethnic conflicts (Horowitz, 2001) or revolutionary movements (Kaplan, 1982) is moderated by the rumors' plausibility.

Some rumors, however, spread even though they are relatively implausible. Many of these rumors spread because they would be very costly to ignore if true. In terms of error management theory, it can make sense to pay attention to, transmit, and potentially endorse beliefs even if they are not entirely plausible (Fessler, Pisor, & Navarrete, 2014). Indeed, both observational (Shibutani, 1966) and experimental (Fessler et al., 2014) studies show this 'negativity bias' to play an important role in rumor transmission. The rationality of paying attention to rumors of dramatic but improbable events is highlighted by historical examples of people who failed to take heed of such rumors (for instance when the US ambassador to Japan rejected a rumor that the Japanese were going to attack Pearl Harbor, Shibutani, 1966, p. 73).

Finally, some rumors spread without being fully believed by people who transmit them, or without playing a significant causal role in their actions. Relatively improbable rumors in particular are often introduced by 'someone told me' or 'people say that' instead of being directly endorsed (Bonhomme, 2009). Quantitative studies suggest that accurate rumors are more likely to be believed than inaccurate ones (DiFonzo & Bordia, 2002). Even rumors that are seemingly believed may only play a post-hoc role. For instance during episodes of inter-ethnic violence, one generally finds inaccurate rumors about the monstrous and threatening behavior of the opposing ethnic group (Horowitz, 2001). These rumors, however, do not seem to cause the resentment that fuels the violence, instead they allow people to coordinate on how to justify breaking social norms (Horowitz, 2001; Turner, 1964).

On the whole, the literature on rumors support Shibutani's (1966, p. 46) conclusion that "[h]uman beings are not gullible." First, because rumors are often accurate. Second, because rumors are mostly consistent with the operation of plausibility checking. Third, because when rumors are implausible, they are either not really believed, or it makes sense for individuals to take them into account from an error management point of view.

2.8 Summary

The second part of this article covers the domains in which accusations of strong gullibility are most often leveled. It suggests that in every case evidence for at least one of the three elements of strong gullibility is wanting—either widespread acceptance of unfounded beliefs, acceptance of costly beliefs, or primacy of source-based processes. For several domains—propaganda, political campaigns, advertising, rumors—a common view among the relevant experts is that the influence of misguided communicated information is generally weak, too weak, at any rate, to fit with strong gullibility. Similarly, historical evidence shows that the extent of the acceptation of world religions by the population at large is often less than suggested by strong gullibility.

Although this claim might be more disputed, I take the evidence to show that communicated beliefs causing people to engage in costly behaviors are even more rarely accepted. For instance, it has been argued that beliefs leading to costly religious behaviors, be they in small-scale societies or in world religions, are relatively rare and are those most likely to be resisted.

Still, there remain many cases in which costly beliefs are accepted on a large scale—for instance, misguided medical beliefs. I have argued that in these cases, it is mostly content-based rather than source-based processes that are at play. Indeed, content-based processes might be responsible for the acceptation of a large majority of empirically unfounded beliefs, from religious beliefs to propaganda or rumors. However, explanations in terms of content-based rather than source-based processes are not consensual, and more research is needed to

further disentangle the role of these two types of processes (see, e.g., Sperber & Claidière, 2008).

3. Conclusion

Section 1 reviews experimental work suggesting that individuals exert epistemic vigilance towards linguistic communication. Individuals pay attention to the plausibility of the message, to the trustworthiness of its source, and to the strength of the arguments supporting the message. On the whole, this evidence shows epistemic vigilance to be efficient in that it relies on sensible cues and enables individuals to reject most misleading messages.

Section 2 reviews a wide range of research in social science bearing on the question of gullibility. More specifically, it tackles strong gullibility—the view that the acceptance of costly communicated beliefs is widespread and largely grounded in source-based processes. This review suggested that most supposed examples of strong gullibility are misinformed. Most persuasion attempts are largely unsuccessful. The efficacy of religious proselytizing, of dictatorial propaganda, of political campaigns, and of advertising is surprisingly limited. When people are influenced by such messages, it is, arguably, mostly because the messages are deemed plausible—that is, thanks to content-based, rather than source-based, processes.

If the strong gullibility view were correct, this would mean that any mechanisms of epistemic vigilance humans might possess are widely dysfunctional. Although the degree of efficacy of epistemic vigilance, both within and outside the confines of the laboratory, has to be further tested, the evidence reviewed here suggests that it is a viable framework for understanding how humans evaluate communicated information.

This conclusion does not mean that there are no exceptions—beliefs that seem to be accepted through source-based processes even though they prove costly. In 1856 and 1857, following the injunction of a prophetess, the Xhosa slaughtered most of their cattle and a devastating famine ensued (Peires, 1989). Some North Koreans seem to be genuinely convinced by parts of Pyongyang's propaganda (e.g. B. R. Myers, 2011), and some Germans kept on liking Hitler after the end of World War II (Kershaw, 1987). Beliefs in moralizing gods sometimes lead to costly behaviors—from the rejection of blood transfusions to martyrdom. It is difficult for children below the age of four to completely mistrust senders (Mascaro & Morin, 2014). And most of us can remember a few instances in which we got tricked into helping someone or buying something for reasons that were, in retrospect, plainly bad.

Even in these cases, however, the hypothesis that we are endowed with well-adapted mechanisms of epistemic vigilance could provide a fruitful framework. As suggested above, our mechanisms of epistemic vigilance likely suffer from a series of adaptive lags. For instance, a salient difference between the environment in which they evolved and our current environment is the length and complexity of the processes by which information reaches us. In small-scale societies, most pieces of information would have only gone through a few individuals before reaching us (e.g. Morin, 2015). By contrast, in our current environment many of our beliefs are the result of long and opaque transmission chains—this is true of historical, scientific, and religious beliefs, and of many beliefs about who we should defer to (Origgi, 2015). There is no reason we should be properly equipped to deal very efficiently with communication stemming from these chains (Mercier & Morin, submitted; Sperber, 2009). Such gaps in epistemic vigilance might explain striking cases of apparent gullibility.

One of the strengths of the epistemic vigilance framework is that it draws attention to the cognitive mechanisms tailored to evaluating communicated information. For instance, it

distinguishes between the operations of source-based and content-based processes (a distinction that is, admittedly, hardly unique to this framework, see, e.g., Boyd & Richerson, 1985). Although both types of processes can lead to the acceptance of misguided communicated beliefs, this acceptation would have different causes and different remedies. For instance, if the spread of some given misguided beliefs—populist political beliefs, say— were chiefly the result of undue deference to a specific source, then attacking this source might be an effective countermeasure. By contrast, if the spread of these beliefs mostly rests on content-based processes, then what is needed is closer to an education effort to convince people to change their preconceived beliefs.

If people are not, on the whole, strongly gullible, what explains the widespread belief to the contrary? Several mechanisms could conspire to make the belief that people are gullible particularly attractive. One has to do with causal attribution. When a given belief is defended by a prestigious figure, and when many people endorse this belief, it is natural to infer that the former caused the latter. In fact the evidence reviewed above suggests that the causality is usually reversed. It is because a creed is widely appealing that those who propound it are successful. That the causality usually runs in this direction is suggested by the fact that when prestigious figures defend unpopular beliefs, they do not manage to spread these beliefs well, and they run the risk of becoming unpopular (even when they can rely on a powerful propaganda apparatus).

A second factor that favors a belief in gullibility is that it can seemingly explain, in a single stroke, why people hold many beliefs we think are false or harmful. For instance, in the U.S. the archetypal Republican disagrees with the archetypal Democrat on many issues—gun control, abortion, redistribution, foreign policy, etc. (these archetypes are in fact rarer in real life than in our imagination, Weeden & Kurzban, 2014). It is thus tempting for each side to explain all of these misguided beliefs by thinking that their opponents gullibly follow party leaders and partisan news networks (moreover, this explanation is not unfounded when it comes to non-attitudes).

A third mechanism is the relative salience of accepting misleading information versus rejecting valuable information. For instance, when we buy a disappointing product, we might blame the ad that pushed us (we think) towards this decision. By contrast, when we miss out on a great product whose ad did not sway us, we might never realize our loss (for a related argument, see, Yamagishi & Yamagishi, 1994). This might lead us to think that gullibility (accepting harmful communicated information) is a more significant issue than conservatism (rejecting beneficial communicated information).

Finally, a fourth factor is the political usefulness of the belief in gullibility. Claiming that people are gullible is a convenient argument against democracy (e.g., Brennan, 2012), as well as freedom of the press and of speech (see Arceneaux & Johnson, 2013). Historically, many of the thinkers who raised the fear of demagogues as an argument against democracy had many other reasons to oppose democracy—such as being part of the dominant class and fearing dispossession or worse.

The fact that people do not seem to be strongly gullible should allay the threat of demagogues. However, it does not constitute a bulletproof argument in favor of democracy, for two reasons. The first is that people are likely to intuitively form wrong beliefs about the best way to run large-scale polities and economies (see, e.g., Caplan, 2007). That these wrongheaded beliefs are recruited by, rather than implanted by demagogues does not make them much less problematic (although it suggests different solutions). The second is that there often are discrepancies between the personal and the societal relevance of our beliefs. For

instance, whether American voters have accurate beliefs about how much the U.S. spends on foreign aid makes little difference to their personal lives. As a result, this belief might be relatively easily influenced by politicians without the voters suffering any cost. However, the beliefs of American voters about foreign aid might be very significant for the potential recipient of this aid. Even if American voters are not gullible, they might still be influenced in societally undesirable ways.

Finally, it should be stressed again that the present article focuses on one of the two ways in which epistemic vigilance can fail: by accepting, rather than rejecting, too much information. Section 1 does provide some elements regarding how epistemic vigilance helps us accept valuable information—for instance how cues to trustworthiness or sound arguments can overcome plausibility checking. But section 2 is almost entirely focused on showing that epistemic vigilance does not err by accepting too much information. To make a case that epistemic vigilance fulfills its function well, one ought to also demonstrate that it allows the transmission of beneficial messages. Given that it likely makes evolutionary sense for epistemic vigilance to err on the side of caution, it is possible that such a study would reveal a fair amount of conservatism. Comparing the relative weight of gullibility and conservatism in explaining communication failures would then make for a theoretically and practically rewarding study.

References

Abercrombie, N., Hill, S., & Turner, B. S. (1980). *The dominant ideology thesis.* London: Allen & Unwin.

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the rise of the nazis in prewar Germany. *The Quarterly Journal of Economics*, *130*(4), 1885–1939.

Akerlof, G. A., & Shiller, R. J. (2015). *Phishing for Phools: The Economics of Manipulation and Deception.* Princeton: Princeton University Press.

Allport, G. W., & Postman, L. (1947). *The psychology of rumor.* Oxford: Henry Holt.

Amos, C., Holmes, G., & Strutton, D. (2008). Exploring the relationship between celebrity endorser effects and advertising effectiveness: A quantitative synthesis of effect size. *International Journal of Advertising*, *27*(2), 209–234.

Anderson, D. E., DePaulo, B. M., & Ansfield, M. E. (2002). The development of deception detection skill: A longitudinal study of same-sex friends. *Personality and Social Psychology Bulletin*, *28*(4), 536–545.

Arceneaux, K., & Johnson, M. (2013). *Changing Minds Or Changing Channels?: Partisan News in an Age of Choice.* Chicago: University of Chicago Press.

Arika, N. (2007). *Passions and Tempers: A History of the Humors.* New York: Harper Perennial.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie Press.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, *70*(9), 1–70.

Baker, S. M., & Petty, R. E. (1994). Majority and minority influence: Source-position imbalance as a determinant of message scrutiny. *Journal of Personality and Social Psychology*, *67*(1), 5.

Barber, B. (1983). *The logic and limits of trust.* New Brunswick, NJ: Rutgers University Press.

Barrett, J. L. (1997). *Anthropomorphism, intentional agents, and conceptualizing God (PhD dissertation)*. Cornell University.

Barrett, J. L. (1999). Theological correctness: Cognitive constraint and the study of religion. *Method & Theory in the Study of Religion*, *11*(4), 325–339.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, *4*(1), 29–34.

Baumard, N., & Boyer, P. (2013a). Explaining moral religions. *Trends in Cognitive Sciences*, *17*(6), 272–280.

Baumard, N., & Boyer, P. (2013b). Religious beliefs as reflective elaborations on intuitions: A modified dual-process model. *Current Directions in Psychological Science*, *22*(4), 295–300.

Baumard, N., & Chevallier, C. (in press). The nature and dynamics of world religions: a life-history approach. In *Proc. R. Soc. B* (Vol. 282, p. 20151593). The Royal Society.

Baumard, N., Hyafil, A., Morris, I., & Boyer, P. (2015). Increased affluence explains the emergence of ascetic wisdoms and moralizing religions. *Current Biology*, *25*(1), 10–15.

Bellah, R. N. (2011). *Religion in human evolution: From the Paleolithic to the Axial Age*. Cambridge: Harvard University Press.

Benjamin, L. T., & Simpson, J. A. (2009). The power of the situation: The impact of Milgram's obedience studies on personality and social psychology. *American Psychologist*, *64*(1), 12.

Bernard, S., Harris, P., Terrier, N., & Clément, F. (2015). Children weigh the number of informants and perceptual uncertainty when identifying objects. *Journal of Experimental Child Psychology*, *136*, 70–81.

Besley, T., Burgess, R., & others. (2002). The Political Economy of Government Responsiveness: Theory and Evidence from India. *The Quarterly Journal of Economics*, *117*(4), 1415–1451.

Blake, T., Nosko, C., & Tadelis, S. (2014). *Consumer heterogeneity and paid search effectiveness: A large scale field experiment*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w20171

Bloch, M. (1989). *Ritual, History and Power Selected Papers in Anthropology*. London: Bloomsbury Academic. Retrieved from https://philpapers.org/rec/BLORHA-3

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *in press*.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214.

Bond, C. F., Howard, A. R., Hutchison, J. L., & Masip, J. (2013). Overlooking the obvious: Incentives to lie. *Basic and Applied Social Psychology*, *35*(2), 212–221.

Bond, R. (2005). Group size and conformity. *Group Processes & Intergroup Relations*, *8*(4), 331–354.

Bonhomme, J. (2009). *Les Voleurs de sexe. Anthropologie d'une rumeur africaine*. Paris: Seuil.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: Chicago University Press.

Boyer, P. (2001). *Religion Explained*. London: Heinemann.

Boyer, P., & Baumard, N. (in press). The diversity of religious systems: An evolutionary and cognitive framework. In J. R. Liddle & T. K. Schackelford (Eds.), *The Oxford Handbook of Evolutionary Perspectives on Religion*. New York: Oxford University Press.

Brandenberger, D. (2012). *Propaganda state in crisis: Soviet ideology, indoctrination, and terror under Stalin, 1927-1941*. New Haven: Yale University Press.

Brennan, J. (2012). *The ethics of voting*. New York: Princeton University Press.

Brock, T. C. (1965). Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology*, *1*(6), 650.

Broockman, D. E., & Green, D. P. (2014). Do online advertisements increase political candidates' name recognition or favorability? Evidence from randomized field experiments. *Political Behavior*, *36*(2), 263–289.

Bullock, J. G. (2011). Elite influence on public opinion in an informed electorate. *American Political Science Review*, *105*(03), 496–515.

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2013). *Partisan bias in factual beliefs about politics*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w19080

Burger, J. M., Girgis, Z. M., & Manning, C. C. (2011). In their own words: Explaining obedience to authority through an examination of participants' comments. *Social Psychological and Personality Science*, *2*, 460–466.

Cacioppo, J. T., & Petty, R. E. (1979). Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of Personality and Social Psychology*, *37*(1), 97–109.

Campbell, J. D., & Fairey, P. J. (1989). Informational and normative routes to conformity: The effect of faction size as a function of norm extremity and attention to the stimulus. *Journal of Personality and Social Psychology*, *57*(3), 457.

Caplan, B. (2007). *The Myth of the Rational Voter*. Princeton: Princeton University Press.

Castelain, T., Bernard, S., Van der Henst, J.-B., & Mercier, H. (2016). The influence of power and reason on young Maya children's endorsement of testimony. *Developmental Science*, *19*(6), 957–966.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752–766.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, *66*(3), 460–473.

Chang, H.-J. (2014). *Economics: The User's Guide*. Bloomsbury Publishing.

Chen, X., & Shi, T. (2001). Media effects on political confidence and trust in the People's Republic of China in the post-Tiananmen period. *East Asia*, *19*(3), 84–118.

Chong, D., & Druckman, J. N. (2007). Framing public opinion in competitive democracies. *American Political Science Review*, *101*(04), 637–655.

Clément, F. (2006). *Les mécanismes de la crédulité*. Geneva: Librairie Droz.

Clement, F., Koenig, M. A., & Harris, P. (2004). The ontogenesis of trust. *Mind and Language*, *19*(4), 360–379.

Clément, F., Koenig, M. A., & Harris, P. (2004). The ontogeny of trust. *Mind and Language*, *19*(4), 360–379.

Cohn, N. (1970). *The Pursuit of the Millennium*. St Albans: Paladin.

Corriveau, K. H., & Harris, P. L. (2010). Preschoolers (sometimes) defer to the majority in making simple perceptual judgments. *Developmental Psychology*, *46*(2), 437.

Coury, C. (1967). The basic principles of medicine in the primitive mind. *Medical History*, *11*(2), 111.

Craig, E. (1999). *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. New York: Oxford University Press.

Davies, S. R. (1997). *Popular opinion in Stalin's Russia: terror, propaganda and dissent, 1934-1941*. Cambridge: Cambridge University Press.

Dawkins, R. (2009). *The god delusion*. New York: Random House.

DellaVigna, S., & Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, *2*(1), 643–669.

DellaVigna, S., & Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *Quarterly Journal of Economics*, *122*.

Delumeau, J. (1977). *Catholicism Between Luther and Voltaire*. Philadelphia: Westminster Press.

Demick, B. (2010). *Nothing to envy: real lives in North Korea*. New York: Spiegel & Grau.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review*, *1*(4), 346–357.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, *74*(1), 63.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995.

Dezecache, G. (2015). Human collective reactions to threat. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 209–219.

Dezecache, G., Mercier, H., & Scott-Phillips, T. C. (2013). An evolutionary approach to emotional communication. *Journal of Pragmatics*, *59*, 221–233.

Dickson, G. (2007). *The Children's Crusade: medieval history, modern mythistory*. London: Palgrave MacMillan.

DiFonzo, N., & Bordia, P. (2002). Corporate rumor activity, belief and accuracy. *Public Relations Review*, *28*(1), 1–19.

DiFonzo, N., & Bordia, P. (2007). *Rumor psychology: Social and organizational approaches.* Washington, DC: American Psychological Association.

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, *19*(6), 695–704.

Druckman, J. N. (2004). Political preference formation: Competition, deliberation, and the (ir) relevance of framing effects. *American Political Science Review*, *98*(04), 671–686.

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, *107*(01), 57–79.

Dumont, L. (1980). *Homo Hierarchicus: The Caste System and Its Implications*. University of Chicago Press.

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, *36*(4), 424.

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition*, *33*(5), 770–782.

Ekelund, R. B., Hébert, R. F., Tollison, R. D., Anderson, G. M., & Davidson, A. B. (1997). *Sacred trust: the medieval church as an economic firm*. New York: Oxford University Press.

Evans-Pritchard, E. E. (1971). *Nuer religion*. Oxford: Oxford University Press.

Femia, J. V. (2001). *Against the Masses: Varieties of Anti-Democratic Thought Since the French Revolution*. New York: Oxford University Press.

Fessler, D. M., Pisor, A. C., & Navarrete, C. D. (2014). Negatively-biased credulity and the cultural evolution of beliefs. *PloS One*, *9*(4), e95167.

Filtzer, D. (1986). *Soviet workers and Stalinist industrialisation*. London: Pluto Press.

Frank, T. (2007). *What's the matter with Kansas?: how conservatives won the heart of America*. New York: Macmillan.

Friedman, E. (2014). *Insurgency trap: Labor politics in postsocialist China*. Ithaca: Cornell University Press.

Friend, R., Rafferty, Y., & Bramel, D. (1990). A puzzling misinterpretation of the Asch "conformity"study. *European Journal of Social Psychology*, *20*(1), 29–44.

Gabbert, F., Memon, A., & Wright, D. B. (2007). I saw it for longer than you: The relationship between perceived encoding duration and memory conformity. *Acta Psychologica*, *124*(3), 319–331.

Gamson, W. A. (1992). *Talking politics*. Cambridge: Cambridge university press.

Gayo-Avello, H. S., Peter Gloor, D., Castillo, C., Mendoza, M., & Poblete, B. (2013). Predicting information credibility in time-sensitive social media. *Internet Research*, *23*(5), 560–588.

Gelman, A., & King, G. (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science*, *23*(04), 409–451.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, *78*(1), 35–71.

Gerard, H. B., Wilhelmy, R. A., & Conolley, E. S. (1968). Conformity and group size. *Journal of Personality and Social Psychology*, *8*(1p1), 79.

Gerber, A. S., Gimpel, J. G., Green, D. P., & Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *American Political Science Review*, *105*(01), 135–150.

Gigerenzer, G. (2007). Fast and frugal heuristics: The tools of bounded rationality. In D. Koehler & N. Harvey (Eds.), *Handbook of Judgment and Decision Making*. Oxford, UK: Blackwell.

Gilbert, D. T. (2006). *Stumbling on Happiness*. New York: Random House.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*(4), 601–613.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*(2), 221–233.

Giner-Sorolila, R., & Chaiken, S. (1997). Selective use of heunrstic and systematic processing under defense motivation. *Personality and Social Psychology Bulletin*, *23*(1), 84–97.

Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74.

Goldman, A. I. (1999). *Knowledge in a social world*. Oxford: Oxford University Press.

Griggs, R. A. (2015). The disappearance of independence in textbook coverage of Asch's social pressure experiments. *Teaching of Psychology*, *42*(2), 137–142.

Griggs, R. A., & Whitehead, G. I. (2015). Coverage of Milgram's Obedience Experiments in Social Psychology Textbooks Where Have All the Criticisms Gone? *Teaching of Psychology*, *42*(4), 315–322.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological Review*, *114*(3), 704–732.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Belknap Press/Harvard University Press.

Harris, P. L., & Lane, J. D. (2013). Infants understand how testimony works. *Topoi*, 1–16.

Hartwig, M., & Bond, C. H. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.

Haslam, S. A., & Reicher, S. D. (2012). Contesting the "nature" of conformity: What Milgram and Zimbardo's studies really show.

Haslam, S. A., Reicher, S. D., & Birney, M. E. (2014). Nothing by mere authority: Evidence that in an experimental analogue of the Milgram paradigm participants are motivated not by orders but by appeals to science. *Journal of Social Issues*, *70*(3), 473–488.

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychological Science*, *16*(7), 566–571.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*(2), 494–50814.

Haun, D. B. M., Rekers, Y., & Tomasello, M. (2014). Children conform to the behavior of peers; other great apes stick with what they know. *Psychological Science*, *25*(12), 2160–2167.

Haun, D. B. M., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development*, *82*(6), 1759–1767.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, *22*(3), 165–196.

Heraclitus. (2001). *Fragments: The collected wisdom of Heraclitus*. (B. Haxton, Trans.). London: Viking Adult.

Herman, E. S., & Chomsky, N. (2008). *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Random House.

Hoeken, H., Šorm, E., & Schellens, P. J. (2014). Arguing about the likelihood of consequences: Laypeople's criteria to distinguish strong arguments from weak ones. *Thinking & Reasoning*, *20*(1), 77–98.

Hoffmann, P. (1991). Internal Resistance in Germany. In *Contending with Hitler* (pp. 119–128). Washington DC: Large.

Holbach, P. H. T. B. d'. (1835). *Christianity Unveiled: Being an Examination of the Principles and Effects of the Christian Religion*. New York: Johnson.

Hollander, M. M. (2015). The repertoire of resistance: Non-compliance with directives in Milgram's "obedience"experiments. *British Journal of Social Psychology*.

Hornikx, J., & Hahn, U. (2012). Reasoning and argumentation: Towards an integrated psychology of argumentation. *Thinking & Reasoning*, *18*(3), 225–243.

Horowitz, D. L. (2001). *The deadly ethnic riot*. Berkeley: University of California Press.

Iyengar, S., & Kinder, D. R. (1987). *News that matters: Television and public opinion*. Chicago: University of Chicago.

Jaeger, A., Lauris, P., Selmeczy, D., & Dobbins, I. G. (2012). The costs and benefits of memory conformity. *Memory & Cognition*, *40*(1), 101–112.

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

Kallis, A. (2008). *Nazi propaganda in the Second World War.* London: Palgrave Macmillan.

Kaplan, S. L. (1982). The famine plot persuasion in eighteenth-century France. *Transactions of the American Philosophical Society*, 1–79.

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications.* Glencoe: Free Press.

Kershaw, I. (1983a). How effective was Nazi propaganda. In D. Welch (Ed.), *Nazi propaganda: the power and the limitations* (pp. 180–205). London: Croom Helm.

Kershaw, I. (1983b). *Popular opinion and political dissent in the Third Reich, Bavaria 1933-1945*. New York: Oxford University Press.

Kershaw, I. (1987). *The Hitler Myth': Image and Reality in the Third Reich*. New York: Oxford University Press.

Kershaw, I. (1991). *Hitler: profiles in power*. London: Routledge.

Kinder, D. R. (1998). Communication and opinion. *Annual Review of Political Science*, *1*(1), 167–197.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294.

Klapper, J. T. (1960). *The Effects of Mass Com- munications*. Glencoe, IL: Free.

Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development*, *83*(3), 1051–1063.

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, *87*(3), 179–208.

Kuklinski, J. H., Luskin, R. C., & Bolland, J. (1991). Where is the schema? Going beyond the "S" word in political psychology. *American Political Science Review*, *85*(04), 1341–1380.

Kuller, C. (2015). The demonstrations in support of the Protestant provincial Bishop Hans Meiser: A successful protest against the Nazi regime. In N. Stoltzfus & B. Maier-Katkin (Eds.), *Protest in Hitler's "National Community": Popular unrest and the Nazi response* (pp. 38–54).

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–15392.

Ladd, J. M. (2011). *Why Americans hate the media and how it matters*. New York: Princeton University Press.

Ladha, K. K. (1992). The Condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science*, 617–634.

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science*, *16*(4), 622–638.

Lane, J. D., & Harris, P. L. (2015). The Roles of Intuition and Informants' Expertise in Children's Epistemic Trust. *Child Development*, *86*(3), 919–926.

Lasswell, H. D. (1972). *Propaganda Technique in the World War*. New York: Garland.

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1948). *The People's Choice. How the Voter Makes up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Le Bras, G. (1955). *Etudes de sociologie religieuse*. Paris: Presses Universitaires de France.

Lenz, G. S. (2009). Learning and opinion change, not priming: Reconsidering the priming hypothesis. *American Journal of Political Science*, *53*(4), 821–837.

Lenz, G. S. (2013). *Follow the leader?: how voters respond to politicians' policies and performance*. Chicago: University of Chicago Press.

Levine, T. R. (2014). Truth-Default Theory (TDT) A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology*, *33*(4), 378–392.

Lippmann, W. (1965). *Public Opinion*. New York: Free Press.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*(4), 560–572.

Luhrmann, T. M. (1991). *Persuasions of the witch's craft: ritual magic in contemporary England*. Cambridge: Harvard University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=mLrK_ZkcxxQC&oi=fnd&pg=PR9&dq=luhrmann+witch%27s+craft&ots=uPRGVlSSKO&sig=6HA3NHJdtgc3gAyDcMcUfFpZyPQ

Lull, R. B., & Bushman, B. J. (2015). Do sex and violence sell? A meta-analytic review of the effects of sexual and violent media and ad content on memory, attitudes, and buying intentions. *Psychological Bulletin*, *141*(5), 1022–1048.

Lupia, A. (1994). Shortcuts versus encyclopedias: information and voting behavior in California insurance reform elections. *American Political Science Review*, *88*(01), 63–76.

Ma, L., & Woolley, J. D. (2013). Young children's sensitivity to speaker gender when learning from others. *Journal of Cognition and Development*, *14*(1), 100–119.

Mackie, D. M., Worth, L. T., & Asuncion, A. G. (1990). Processing of persuasive in-group messages. *Journal of Personality and Social Psychology*, *58*(5), 812.

Maillat, D., & Oswald, S. (2009). Defining manipulative discourse: The pragmatics of cognitive illusions. *International Review of Pragmatics*, *1*(2), 348–370.

March, C., Krügel, S., & Ziegelmeyer, A. (2012). Do We Follow Private Information when We Should? Laboratory Evidence on Naıve Herding. *Jena Economic Research Papers*, 002.

Marx, K., & Engels, F. (1970). *The German Ideology*. International Publishers Co.

Mascaro, O., & Morin, O. (2014). Gullible's travel: How honest and trustful children become vigilant communicators. In L. Robinson & S. Einav (Eds.), *Trust and Skepticism: Children's Selective Learning From Testimony*. London: Psychology Press.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*, 367–380.

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford: Oxford University Press.

McElreath, R., Lubell, M., Richerson, P. J., Waring, T. M., Baum, W., Edsten, E., … Paciotti, B. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior*, *26*(6), 483–508.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*(5), 469.

Mercier, H., Bernard, S., & Clément, F. (2014). Early sensitivity to arguments: How preschoolers weight circular arguments. *Journal of Experimental Child Psychology*, *125*, 102–109.

Mercier, H., & Morin, O. (submitted). Informational conformity: how good are we at aggregating convergent opinions?

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74.

Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York: Harper & Row.

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, *49*(3), 404.

Miton, H., Claidière, N., & Mercier, H. (2015). Universal cognitive mechanisms explain the cultural success of bloodletting. *Evolution and Human Behavior*, *36*(4), 303–312.

Miton, H., & Mercier, H. (2015). Cognitive obstacles to pro-vaccination beliefs. *Trends In Cognitive Sciences*, *19*(11), 633–636.

Morgan, T. J. H., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: effects of uncertainty and consensus. *Developmental Science*, *18*(4), 511–524.

Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1729), 653–662.

Morin, O. (2015). *How Traditions Live and Die*. New York: Oxford University Press.

Moscovici, S. (1976). *Social influence and social change* (Vol. 10). London: Academic Press London.

Murray, A. (1974). Religion among the poor in thirteenth-century France: the testimony of Humbert de Romans. *Traditio*, 285–324.

Mutz, D. C. (1998). *Impersonal influence: How perceptions of mass collectives affect political attitudes*. Cambridge: Cambridge University Press.

Myers, B. R. (2011). *The cleanest race: how North Koreans see themselves and why it matters*. Brooklyn: Melville House.

Myers, D. G. (2011). *Psychology*. New York: Worth.

Ober, J. (1993). Thucydides' Criticism of Democratic Knowledge. *Nomodeiktes: Greek Studies in Honor of Martin Ostwald*, 81–98.

O'Brien, K. J., & Li, L. (2008). *Rightful resistance in rural China*. Cambridge: Cambridge University Press.

Oliver, R. L. (2014). *Satisfaction: A behavioral perspective on the consumer*. London: Routledge.

Origgi, G. (2015). *La réputation*. Paris: Presses Universitaires de France.

Peires, J. B. (1989). *The dead will arise: Nongqawuse and the great Xhosa cattle-killing movement of 1856-7*. Bloomington: Indiana University Press.

Perry, G. (2013). *Behind the shock machine: The untold story of the notorious Milgram psychology experiments*. New York: The New Press.

Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology*, *37*, 349–360.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847–855.

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 323–390). Boston: McGraw-Hill.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–784.

Poulantzas, N. (1978). *Political Power and Social Classes*. London: Verso.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*, 39–57.

Prior, M., & Lupia, A. (2008). Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science*, *52*(1), 169–183.

Puckett, J. M., Petty, R. E., Cacioppo, J. T., & Fischer, D. L. (1983). The relative impact of age and attractiveness stereotypes on persuasion. *Journal of Gerontology*, *38*(3), 340–343.

Reid, T. (2000). *An Inquiry Into The Human Mind: On The Principles of Common Sense*. Edinburgh: Edinburgh University Press.

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, *96*, 538–558.

Rips, L. J. (1998). Reasoning and conversation. *Psychological Review*, *105*, 411–441.

Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer uterrance veracity from speaker informedness. *Developmental Psychology*, *35*(2), 535–546.

Ryckman, R. M., Rodda, W. C., & Sherman, M. F. (1972). Locus of control and expertise relevance as determinants of changes in opinion about student activism. *The Journal of Social Psychology*, *88*(1), 107–114.

Salter, S. (1983). Structures of consensus and coercion: Workers morale and the maintenance of work discipline, 1939-1945. In D. Welch (Ed.), *Nazi propaganda: the power and the limitations* (pp. 88–116). London: Croom Helm.

Schlenker, B. R., Helm, B., & Tedeschi, J. T. (1973). The effects of personality and situational variables on behavioral trust. *Journal of Personality and Social Psychology*, *25*(3), 419–427.

Schweingruber, D., & Wohlstein, R. T. (2005). The madding crowd goes to school: Myths about crowds in introductory sociology textbooks. *Teaching Sociology*, *33*(2), 136–153.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, *101*(1), 1–19.

Scott, J. C. (1990). *Domination and the arts of resistance: Hidden transcripts*. New Haven: Yale university press.

Scott-Phillips, T. C. (2014). *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. London: Palgrave MacMillan.

Sethuraman, R., Tellis, G. J., & Briesch, R. A. (2011). How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities. *Journal of Marketing Research*, *48*(3), 457–471.

Shibutani, T. (1966). *Improvised News. A Sociological Study of Rumor*. New York: Bobbs-Merrill Company.

Signer, M. (2009). *Demagogue: the fight to save democracy from its worst enemies*. New York: Macmillan.

Slone, J. (2007). *Theological Incorrectness: Why Religious People Believe What They Shouldn't*. New York: Oxford University Press.

Snyder, J. M., & Strömberg, D. (2010). Press Coverage and Political Accountability. *Journal of Political Economy*, *118*(2), 355–408.

Sperber, D. (1975). *Rethinking Symbolism*. Cambridge: Cambridge University Press.

Sperber, D. (1994). Understanding verbal understanding. In J. Khalfa (Ed.), *What is Intelligence?* (pp. 179–198). Cambridge, MA: Cambridge University Press.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind and Language*, *12*(1), 67–83.

Sperber, D. (2009). Culturally transmitted misbeliefs. *Behavioral and Brain Sciences*, *32*(6), 534.

Sperber, D., & Claidière, N. (2008). Defining and explaining culture (comments on Richerson and Boyd, Not by genes alone). *Biology & Philosophy*, *23*(2), 283–292.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind and Language*, *25*(4), 359–393.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. New York: Wiley-Blackwell.

Spry, A., Pappu, R., & Bettina Cornwell, T. (2011). Celebrity endorsement, brand credibility and brand equity. *European Journal of Marketing*, *45*(6), 882–909.

Stanley, J. (2015). *How Propaganda Works*. New York: Princeton University Press.

Stark, R. (1999). Secularization, rip. *Sociology of Religion*, *60*(3), 249–273.

Stimson, J. A. (2004). *Tides of consent: How public opinion shapes American politics*. Cambridge: Cambridge University Press.

Street, C. N. H. (in press). ALIED: Humans as adaptive lie detectors. *Journal of Applied Research in Memory and Cognition*.

Street, C. N. H., & Richardson, D. C. (2015a). Descartes versus Spinoza: Truth, uncertainty, and bias. *Social Cognition*, *33*, 1–12.

Street, C. N. H., & Richardson, D. C. (2015b). Lies, Damn Lies, and Expectations: How Base Rates Inform Lie–Truth Judgments. *Applied Cognitive Psychology*, *29*(1), 149–155.

Strömberg, D. (2004). Radio's impact on public spending. *The Quarterly Journal of Economics*, 189–221.

Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granwag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 229–250). New York: Cambridge University Press.

Taylor, M. G. (2013). Gender influences on children's selective trust of adult testimony. *Journal of Experimental Child Psychology*, *115*(4), 672–690.

Teiwes, F. C. (1984). *Leadership, legitimacy, and conflict in China: from a charismatic Mao to the politics of succession*. London: Routledge.

Tellis, G. J. (2003). *Effective advertising: Understanding when, how, and why advertising works*. London: Sage Publications.

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science*, *18*(1), 46–50.

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, *44*(5), 1368–1375.

Terrier, N., Bernard, S., Mercier, H., & Clément, F. (submitted). Visual Access Trumps Gender in 3- and 4-year-old Children's Endorsement of Testimony.

Thomas, K. (1971). *Religion and the Decline of Magic*. London: Weidenfeld and Nicolson.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted mind: evolutionary psychology and the generation of culture* (p. 19).

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: the evolutionary roots of organizational behavior. *Managerial and Decision Economics*, *27*(2–3), 103–129.

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, *143*(5), 1958–1971.

Trouche, E., Shao, J., & Mercier, H. (submitted). How is argument evaluation biased?

Tulis, J. (1987). *The Rhetorical Presidency*. New York: Princeton University Press.

Turner, R. H. (1964). Collective behavior. In R. E. L. Paris (Ed.), *Handbook of modern sociology* (pp. 382–425). Chicago: Rand McNally.

Van Den Putte, B. (2009). What matters most in advertising campaigns? The relative effect of media expenditure and message content strategy. *International Journal of Advertising*, *28*(4), 669–690.

Voigtländer, N., & Voth, H.-J. (2015). Nazi indoctrination and anti-Semitic beliefs in Germany. *Proceedings of the National Academy of Sciences*, *112*(26), 7931–7936.

Vullioud, C., Clément, F., Scott-Phillips, T. C., & Mercier, H. (submitted). Confidence as an expression of commitment: Why overconfidence backfires.

Wang, S. (1995). *Failure of Charisma: The Cultural Revolution in Wuhan*. New York: Oxford University Press.

Weber, M. (1958). *From Max Weber: Essays in Sociology*. New York: Oxford University Press.

Weeden, J., & Kurzban, R. (2014). *The hidden agenda of the political mind: How self-interest shapes our opinions and why we won't admit it*. Princeton: Princeton University Press.

Weizsäcker, G. (2010). Do We Follow Others When We Should? A Simple Test of Rational Expectations. *American Economic Review*, *100*(5), 2340–60.

Williams, B. (2002). *Truth and Truthfulness: An Essay in Genealogy*. New York: Princeton University Press.

Wootton, D. (2006). *Bad medicine: doctors doing harm since Hippocrates*. Oxford: Oxford University Press.

Worsley, P. (1957). *The trumpet shall sound: A study of" cargo" cults in Melanesia*. London: MacGibbon & Kee.

Wright, R. (2009). *The Evolution of God*. New York: Little, Brown.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 129–166.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*, 1–13.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260–281.

Zimbardo, P. G., Johnson, R. L., & McCann, V. (2006). *Psychology: core concepts*. Boston: Pearson.