PNAS INAUGURAL ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Belief traps: Tackling the inertia of harmful beliefs

Marten Scheffer[a,1], Denny Borsboom[b], Sander Nieuwenhuis[c], and Frances Westley[d]

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2019.
Contributed by Marten Scheffer; received February 21, 2022; accepted June 6, 2022; reviewed by Kenneth Kendler and Leor Zmigrod

Beliefs can be highly resilient in the sense that they are not easily abandoned in the face of counterevidence. This has the advantage of guiding consistent behavior and judgments but may also have destructive consequences for individuals, nature, and society. For instance, pathological beliefs can sustain psychiatric disorders, the belief that rhinoceros horn is an aphrodisiac may drive a species extinct, beliefs about gender or race may fuel discrimination, and belief in conspiracy theories can undermine democracy. Here, we present a unifying framework of how self-amplifying feedbacks shape the inertia of beliefs on levels ranging from neuronal networks to social systems. Sustained exposure to counterevidence can destabilize rigid beliefs but requires organized rational override as in cognitive behavioral therapy for pathological beliefs or institutional control of discrimination to reduce racial biases. Black-and-white thinking is a major risk factor for the formation of resilient beliefs associated with psychiatric disorders as well as prejudices and conspiracy thinking. Such dichotomous thinking is characteristic of a lack of cognitive resources, which may be exacerbated by stress. This could help explain why conspiracy thinking and psychiatric disorders tend to peak during crises. A corollary is that addressing social factors such as poverty, social cleavage, and lack of education may be the most effective way to prevent the emergence of rigid beliefs, and thus of problems ranging from psychiatric disorders to prejudices, conspiracy theories, and posttruth politics.

beliefs | psychiatric disorders | conspiracy thinking | inequality | education

Significance

Beliefs are a key element of healthy cognition. Yet overly rigid beliefs are the basis of societal problems including prejudices, psychiatric disorders, and conspiracy theories. Recent findings from neurobiology, psychiatry, and social sciences show how resilience of beliefs is boosted by stressful conditions. This implies the possibility of self-propelled societal deterioration where rigid beliefs harm the quality of personal and political decisions, evoking more-stressful conditions that further rigidify beliefs. Measures reducing social stress, including economic policies such as universal base income, may be the most effective ways to counteract this vicious cycle.

A belief is something that a person holds to be true. Without beliefs, we cannot function. Think of beliefs such as "today is Friday," "bread is edible," or "the keys are in my pocket." Beliefs are thus an essential feature of the way our mind works, and their nature has been discussed for centuries by philosophers (1, 2). At the same time, stubborn beliefs can be a real problem. Indeed, thinkers from Spinoza (3) to Pinker (4) have pointed out that dogmatic belief is among the main hurdles for human progress. The capacity to form overly rigid beliefs seems to be hardwired in human nature, sometimes with major negative consequences. On an individual level, false beliefs may lead to unwise decisions, and pathological forms may cause untold suffering. On a societal level, unfounded beliefs may invoke behavior that has enormous costs. For instance, the belief that rhinoceros horn works as an aphrodisiac is about to drive emblematic species extinct (5). Similarly, beliefs about albino people make their lives highly uncertain in Africa (6), and beliefs about witches had disastrous outcomes for many people in history (7). On a more subtle level, implicit beliefs about intrinsic capacities related to gender or race frustrate discriminated groups, perpetuate inequalities, and imply underutilization of human potential. Also, rigid yet false beliefs in dangerous side effects of vaccinations may reduce vaccination rates, exposing society to risks of dangerous epidemics, and beliefs in conspiracy theories may hamper the functioning of democracies.

This raises the question of what society can do to reduce the inertia of problematic beliefs. The first solution that comes to mind is that people should be exposed to better information. However, while information may help to destabilize beliefs (8), many beliefs are quite irresponsive to counterevidence. For instance, exposure to results of a large metaanalysis concluding that organic food does not offer significant nutritional advantages did not affect the opinion of proorganic readers (9). Indeed, a lack of responsiveness to evidence may be seen as an inherent property of beliefs. Ironically, too much faith in the power of evidence in shaping attitudes is an example of the very same fallacy. Most of us overestimate the role of rational thinking in determining our behavior and attitudes (10, 11). Scientists are no exception to this bias, as illustrated by our astonishment about the prevailing neglect of evidence in decision-making. Surely, as scientists, we should aim to suppress our own rational fallacy, accept that evidence plays a minor role in shaping beliefs, and ask how we can advise society better on ways to tackle the inertia of unwanted beliefs.

Author affiliations: [a]Department of Ecology and Evolution, Wageningen University & Research, 6700 AA Wageningen, The Netherlands; [b]Universiteit van Amsterdam, 1012 WX Amsterdam, The Netherlands; [c]Universiteit Leiden, 2311 EZ Leiden, The Netherlands; and [d]University of Waterloo, Waterloo, ON N2L3G1, Canada

Author contributions: M.S., D.B., S.N., and F.W. wrote the paper.

Reviewers: K.K., Virginia Commonwealth University; and L.Z., University of Cambridge.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: marten.scheffer@wur.nl.

Published July 18, 2022.

Perhaps the main reason that there is no coherent understanding of the inertia of beliefs is that mechanisms that cause beliefs to be resilient act on very different organizational levels, and thus fall in the realm of different branches of science. Neuroscientists are studying the fundamental mechanisms that explain why the brain tends to lock into distinct interpretations and stick to them, a phenomenon illustrated in a simple way by the perception of ambiguous images such as the classical rabbit–duck illusion (Fig. 1). At the same time, psychologists are starting to understand how the resilience of pathological beliefs ("spiders are dangerous," "I am worthless," etc.) can be overcome by therapies (12), and how character traits (e.g., intellectual humility) that affect a person's capacity to avoid belief traps link to perceptual and cognitive functions (13). Importantly, many beliefs are also shaped by social processes. Most of us hold largely the same views as our friends and family, especially on complex issues where our own observations are insufficient to form an opinion. As social scientists have shown, sharing beliefs is important to feel part of a group, and mutual "contagion" can cause such shared beliefs to have strong inertia (14, 15). Social media have taken this lock-in effect to a next level, as sharing news and views among friends in "echo chambers" tends to reduce exposure to belief-challenging information, resulting in more-extreme attitudes over time (16).

In this perspective, we review the fundamental mechanisms that cause inertia of beliefs. We first address the well-studied tendency to preferentially attend to evidence for (rather than against) beliefs we hold, and use that phenomenon to develop a graphical model illustrating how we may view inertia of beliefs in terms of the theory of resilience and tipping points. Subsequently, we look at the way in which coherence within networks of beliefs and social feedback loops may reinforce resilience, and at the neural mechanisms ultimately coding beliefs. Against this background, we ask how resilience of unwanted beliefs may be destabilized in practice, focusing on two classes of examples: pathological beliefs that shape mental disorders and socially embedded beliefs that fuel problematic behavior such as discrimination, vaccine refusal, and violence.

How Beliefs Become Resilient against Counterevidence

Beliefs as Attractors. Despite their apparent inertia, beliefs are shaped and maintained by an ongoing interplay of dynamical processes. Therefore, one obvious way to understand alternative beliefs is to see them as "attractors" in the sense of dynamical systems theory. The generic condition for alternative attractors to occur is the existence of self-propelling feedbacks (18).

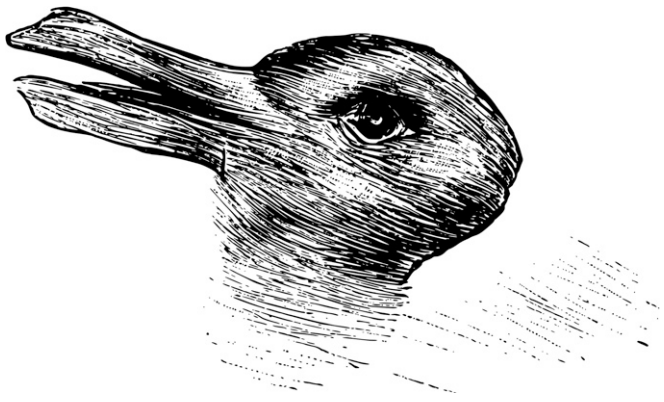


Fig. 1. The rabbit–duck illusion from the 23 October 1892 issue of the German magazine *Fliegende Blätter* (17).

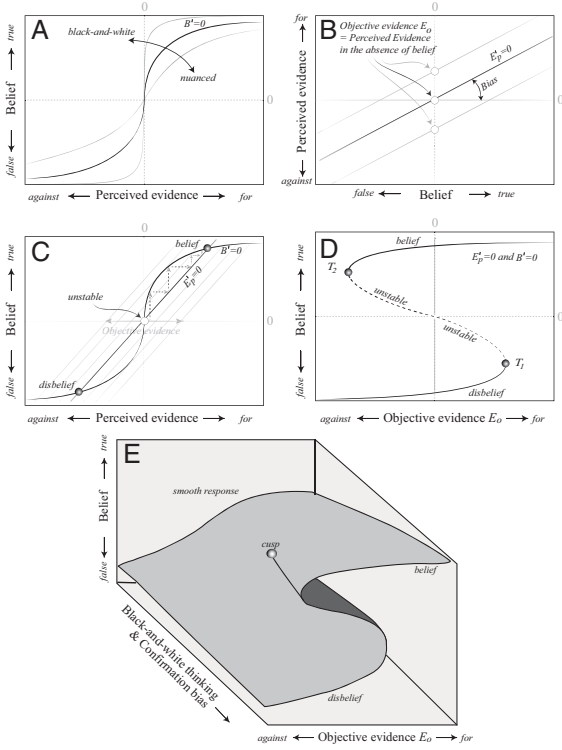
At the level of the individual, an obvious mechanism that may stabilize beliefs is confirmation bias (19). As Francis Bacon wrote almost four centuries ago (20): "The human understanding when it has once adopted an opinion … draws all things else to support and agree with it." This obviously tends to deepen beliefs over time, and also explains the disproportionate effect of a first impression, known as the primacy effect. Quoting Bacon again, "the first impression colors and brings into conformity with itself all that come after." Countless experimental studies have since demonstrated the powerful effects of confirmation bias in everyday life as well as in professional fields such as medicine, justice, and science (19).

Although it makes intuitive sense that self-amplification can cause beliefs to become traps from which escape is hard, such qualitative narratives do not allow us to see how beliefs may nonetheless become fragile to the point that a single additional observation tips the balance, causing the belief to be abandoned. Clearly, not all beliefs are equally resilient, and the resilience of a belief can change over time. Understanding resilience of beliefs is, in a sense, the holy grail if we agree with Bacon, Spinoza, and Pinker that beliefs are one of the main impediments to human progress. If we want to do something about that, we should find ways to reduce the resilience of beliefs. But how does this work? How does resilience of a belief emerge from the continuous interplay of dynamical processes?

To see how we may think of resilience of a belief, consider a simple graphical model of confirmation bias (Box 1). The model describes the dynamic interaction between the strength of a belief and the perceived evidence, assuming that the strength of a belief increases with the perceived evidence (panel *A*) while, at the same time, the strength of the perceived evidence is amplified by the strength of the belief (panel *B*). It can be easily seen how this dynamic interplay can cause belief and disbelief to be alternative stable states (panel *C*). Both states can persist over a wide range of objective evidence (between T_1 and T_2 in panel *D*). However, while both states are stable over that range, their resilience changes, depending on the evidence. One way to see, intuitively, what this implies in terms of stability and resilience is to plot "stability landscapes" (or "potential landscapes") for different levels of objective evidence (Fig. 2). Close to the tipping points T_1 and T_2, the basin of attraction (the valley) around the corresponding attractors becomes small, implying a loss of resilience. For instance, with increasing evidence-against, the basin of attraction around the belief attractor shrinks, implying that the belief becomes fragile in the sense that a small perturbation to the belief, such as a conversation with a nonbelieving friend, may tip the system into the attraction basin of the disbelief attractor.

This example thus illustrates how a sufficient amount of evidence may eventually destabilize a belief or disbelief. Indeed, even notorious conspiracy beliefs can be destabilized by exposure to counterevidence, especially when such evidence is presented by a source that is perceived as trustworthy (8). Yet, evidence may often have surprisingly little impact. In terms of our model, this implies that there is a substantial hysteresis effect, meaning that there is a large range of objective evidence between the two tipping points (T_1 and T_2 in Box 1 and Fig. 2) for which the belief attractor and the disbelief attractor coexist. Solving the big question of how we may reduce the resilience of beliefs, lamented by Spinoza and Pinker, thus essentially requires understanding what determines this hysteresis effect. Our model suggests that hysteresis is larger if the response to perceived evidence

Box 1. The Effect of Confirmation Bias on Belief Resilience, a Graphical Model



A graphical model of the dynamic interaction between the strength of a belief and the perceived evidence. We assume that belief is a saturating function of the perceived evidence-for, and that disbelief (negative belief) saturates with perceived evidence-against. The result (*A*) is a sigmoidal curve representing belief strength as a function of perceived cumulative evidence. Note that an extremely black-and-white thinking subject might flip from complete disbelief to complete belief around the perceived evidence of zero, whereas, in nuanced persons, we would see a gradual change of the strength of the belief with perceived evidence resulting in a smoother sigmoid. The crux of the confirmation bias is that perceived evidence is a function not only of the objective evidence but also of the strength of the belief. Thus, depending on the strength of the confirmation bias, an objectively neutral package of evidence might be turned into evidence for or against, depending on the existing belief (*B*, solid line; note that the axes are flipped to make it easier to see belief as a driver of perceived evidence). Similarly, objective evidence-against might still be seen as evidence-for if the belief is strong enough (*B*, lower gray line, right-hand side). The lines in *A* and *B* can be interpreted as showing the equilibrium of belief strength as a function of perceived evidence ($B' = 0$ in *A*), and the equilibrium of perceived evidence as a function of belief strength ($E'_p = 0$ in *B*). To see the equilibria of the interactive dynamics of belief and perceived evidence together, we combine the two in one plot (*C*). Here, intersections represent equilibria of the system as a whole, where both belief strength and perceived evidence are in equilibrium. The middle intersection represents an unstable equilibrium, also known as a repellor. To see why, imagine starting perturbing it with a tiny bit of evidence, and find the equilibrium belief corresponding to that (lower dashed arrow). From there, find the new perceived evidence strength corresponding to that new belief level (next dashed arrow), and so on. This zig-zag path propels the system toward the "belief" equilibrium (upper solid dot). Starting from different points in the graph, this simple exercise shows that the system is attracted to either the "belief" equilibrium or to the alternative "disbelief" one (lower left dot). Thus, this minimal model has two attractors and one repellor. The neutral (zero-belief) state is unstable, as any observation will trigger a belief which colors the observation, reinforcing the belief and so forth until an equilibrium of pronounced belief or disbelief is reached. Now imagine what will happen if the objective evidence is changing. In our model, this implies that the equilibrium line of perceived evidence ($E'_p = 0$) is shifting (gray lines in *C*). As a result, the positions of the intersection points with the sigmoidal belief curve ($B' = 0$) will also change. As those intersection points are the equilibria of the interactive system, we can see the effect of objective evidence (E_0) on the systems equilibria, by plotting the equilibrium belief strength as a function of objective evidence (*D*). The resulting folded curve has two stable parts, corresponding to the "belief" and the "disbelief" attractors, separated by an unstable part corresponding to the repellor. Thus, even if objective evidence changes, this will usually only have minor effects on the perceived evidence and, therefore, on the belief. However, if the cumulative objective evidence changes strongly enough for the unstable equilibrium to touch the belief equilibrium, a tipping point (T_2) is reached where stability of the belief is destroyed, and the belief will be abandoned. Analogously, in T_1, the disbelief becomes unstable. The tendency to get stuck in alternative attractors is what we call hysteresis, in dynamical systems theory. Note that hysteresis happens only if line ($E'_p = 0$) and sigmoid ($B' = 0$) can have multiple intersections. This requires the maximum slope of the sigmoid ($B' = 0$) to be steeper than the slope of the line ($E'_p = 0$) and thus a sufficiently strong inclination to black-and-white thinking and confirmation bias (note the swapped axes relative to *B*). Depending on those inclinations, subjects thus may respond smoothly the cumulative evidence at hand, or tend become trapped in belief or disbelief (*E*).

is more "black and white" (panel A of the figure in box 1) and if confirmation bias is stronger, so that evidence becomes more "colored" by the belief itself ($E'_p = 0$ is less horizontal in Box 1, panel *B* or less vertical in Box 1, panel *C*). This implies that reducing either black-and-white thinking or the strength of confirmation bias would help. Both may be dampened through a "rational override" of our intuitive responses, or, as Nobel laureate Daniel Kahneman (11) might frame it, strengthening "slow thinking" to correct the biases in our "fast thinking." Also, the tendency for black-and-white thinking may be diminished by reducing stress. We will return to the practical question of how societies may reduce the inertia of harmful beliefs later.

Stabilization of Beliefs by Cognitive Webs and Social Networks. While this model illustrates the basic principles of resilience and hysteresis, it is, of course, a very crude simplification of how the dynamic interplay between a belief and confirmation bias might work. Moreover, the feedback generated by confirmation bias is just one of the mechanisms that contribute to stabilizing beliefs. There are two higher organizational levels at which isolated beliefs can be stabilized. First, in the individual brain, the tendency to strive for an internally consistent worldview may promote inertia of beliefs (21–23). As classical studies have shown (23), we tend to avoid "cognitive dissonance," a term for perceived logical incoherence of views one holds. The resulting psychological discomfort is a strong motivator to resolve such lack of coherence. This often leads to rejection of counterevidence for a held belief, with sometimes surprising effects. For instance, in one study, dire warnings of the dangers of global warming increased skepticism about global warming in individuals believing that the world is fundamentally just (24). Rather than abandoning their "just world" view, subjects resolved the cognitive dissonance (23) by rejecting information that challenged it, effectively concluding that global warming does not exist.

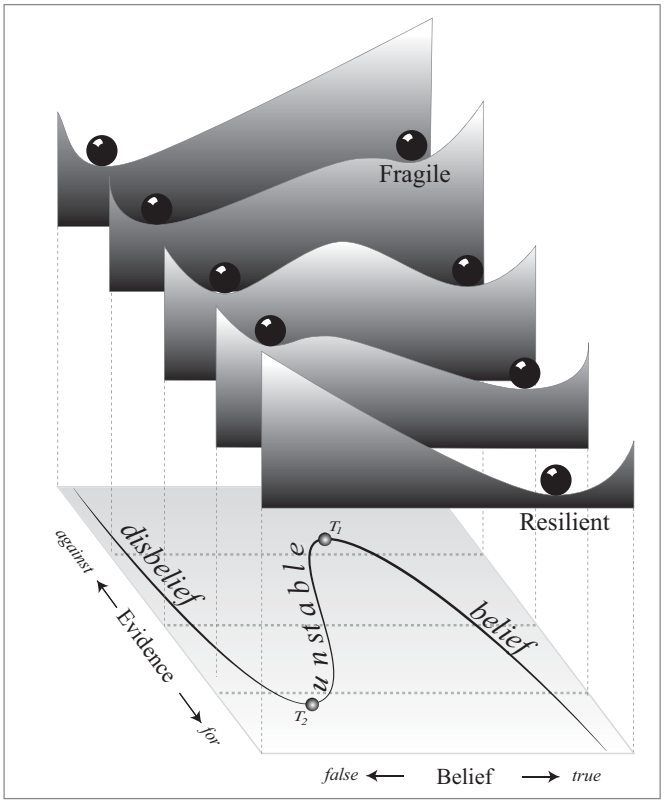


Fig. 2. Stability landscapes illustrating how the resilience of a belief may change as a function of objective evidence. The slope of such landscapes represents the speed and direction of change. A zero slope (horizontal) means no change, thus an equilibrium. The horizontal top of the hill is an unstable equilibrium, as the tiniest perturbation will cause the system, visualized by the ball, to roll to the belief or disbelief attractors. Those attractors are at the bottom of valleys, implying that they are stable equilibria to which the ball will roll back upon small perturbations. Note that the sigmoidal equilibrium curve on the bottom plane corresponds to the curve in panel *D* of Box 1.

The second organizational level on which beliefs can be stabilized is social. Our worldviews are stabilized by the fact that we are part of a network of people that tend to hold similar worldviews (25). This is because "birds of a feather flock together" (26) but also due to the continuous process of mutual "contagion" (27). If I tend to believe what you believe, and you tend to believe what I believe, the resulting feedback loop promotes the resilience of the belief. This can be generalized to larger groups (28). One can imagine that there is a social cost to holding a minority belief. As more peers hold a particular belief, the cost of deviating goes up, driving even more peers to hold the same belief, and so on. This amplifying feedback can cause inertia in the sense that strong cumulative counterevidence may be needed to shift a social group away from a currently held belief or attitude (15, 27). The flip side of the coin is that, once a shift away from the belief begins, it is again self-amplifying. For instance, as more and more peers become disbelievers, it becomes socially unattractive to stay behind.

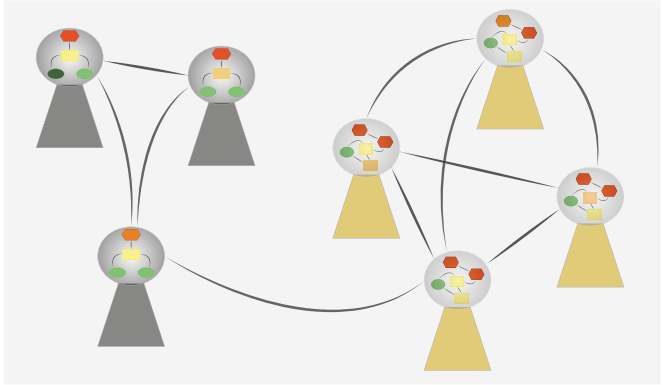


Fig. 3. Alternative ideologies as networks of coherent beliefs embedded in social networks. Each of the colored elements represents an individual belief, linked to other beliefs within the same person in a coherent network representing an ideology. People are attracted to social networks that hold a similar ideology, and contagion within such networks promotes further convergence to the same ideology. Adapted from ref. 25.

The literatures on mechanisms that shape coherence and social dynamics of attitudes are vast (25, 28, 29), and extending the minimal model we presented to describe the resilience of belief networks embedded in social networks would go too far in this context. However, looking from the point of view of any single belief (one of the colored elements in Fig. 3), it is obvious that its resilience may be promoted by coherence (i.e., lack of dissonance) with the rest of the network of beliefs (the ideology) that a person holds, and by the prevalence of that belief in the social network to which the person belongs. Framed otherwise, abandoning a belief can have high costs if it implies destabilization of one's entire worldview, and more so if this would challenge one's membership of a social group ("social identity") (29).

An influential early case study is described in the classical book *When Prophecy Fails* (30). The authors studied a small cult around a woman who claimed to receive messages from another planet, including the prophecy that large portions of the world would be destroyed by a flood on December 21, 1954. Dedicated members left their jobs, ended relationships with nonbelievers, and gave away possessions in preparation for departure on a flying saucer that would rescue them. As the prophesied date passed without signs of a flood, surprisingly many of the members, rather than abandoning the belief, became even more committed, developing various rationalizations for the absence of the flood. Thus, they dissolved the cognitive dissonance (31) that arises from conflicting evidence, by reinterpreting the evidence rather than by updating their beliefs. While this study and other classical work on shared beliefs emphasizes movements where people really meet, social coherence in virtual communities is also shaped by commonly held beliefs, and longing to be part of the community is a powerful driver, as illustrated by a moving report from a *New York Times* journalist listening in to a QAnon group for 3 wk (32). We will return to these social dimensions of beliefs when we address the practical question of how resilience of beliefs may be reduced.

The Neural Basis of Belief Resilience. The mechanisms that shape beliefs must be physically realized in the brain somehow. This raises the question of how resilience of beliefs may be understood on a neurobiological level. Are there really attractors and tipping points associated with the physical representations of beliefs? For simple perceptual beliefs such as the recognition of a person or an object, there is, indeed, good evidence that attractor dynamics play an important role (33, 34). Much of the literature on this topic is rather technical, but the following simplified story may illustrate the core ideas. Imagine you are walking on a long empty beach and see something far off on the sand. What is it? Is it a dead bird, a rock, a fish? One can think of bird, rock, and fish as concepts, each corresponding to an attractor state of the neural network. Together with other concepts, we could represent them as a landscape with many valleys, reflecting basins of attraction. As you are moving closer along the beach, the perceptual evidence needed to identify the object steadily

accumulates. Suddenly, "the penny drops," and you see it: Somebody left a handbag! We all know that experience when you suddenly "see" something—it can be the solution to a complex problem, or just the sudden recognition of someone or something. As the penny drops, your brain shifts from a labile input-sensitive state into a stable attractor state where it is much less affected by subsequent inputs (35, 36). This attractor state reflects the brain's commitment to one categorical interpretation of the incoming sensory data (it's a handbag) (35–37).

Early ideas (33, 38) about how object recognition and decision-making play out in a landscape of alternative neuronal attractors are now being supported by neuroscientific findings. The simple perceptual beliefs on which most of the experimental work has focused do not correspond directly to complex longer-term beliefs (e.g., the planet's spinning in the solar system causes the succession of day and night). Nonetheless, neurobiological measurements have revealed abrupt and widespread shifts in neural activity in the prefrontal cortex as animals adjust their "world view" in response to changes in their experimentally controlled environment, such as altered task–reward rules (34, 39). The dropping of the penny has also been visualized in the brain activity of human subjects using EEG: Looking at an image that gradually became less vague, the subjects' EEG activity exhibited a characteristic abrupt wave, known as the P300, that coincided with the sudden recognition of the image (36, 37).

As brain activity converges toward one of the attractor states representing the alternative beliefs, neural activity shows a rapid decrease in variability and a quickly diminishing influence of newly arriving information (40, 41). This finding is consistent with another hallmark of attractor states: low variability as the system reaches the deep basin of a resilient attractor (42), in this case, corresponding to strong commitment to a decision, belief, or action. As we will discuss later, the presence of such resilient states is important, as it allows the brain to maintain short-term memories that are resistant to distraction.

The properties of the brain that produce these attractor dynamics during simple decision-making tasks are relatively well understood. Experimental and computational work (35, 43–45) has led to a model in which each choice alternative (i.e., belief about the world) is represented by a separate pool of neurons. Recurrent excitatory connections between the neurons in each pool sustain activity triggered by a stimulus. At the same time, the different pools of neurons compete with each other through inhibition. Upon a stimulus, the firing rates of competing pools of neurons initially increase together (because of stimulus ambiguity or noise), but, at some point, the activity of one of the pools suddenly rises (due to accumulating evidence in favor of one choice) while the others are rapidly suppressed (due to winner-take-all competition caused by the feedback inhibition). The choice is thus determined by which of the alternative attractors wins the competition and reaches a stable attractor state. Interestingly, these competitive dynamics are amplified during states of increased arousal or stress (46, 47). This has the advantage that competition between attractors is resolved faster, but also implies that existing biases are magnified. This neurobiological finding resonates with the observation that stress promotes black-and-white thinking (48, 49).

The essence of our confirmation bias model (Box 1) is also supported by experiments studying the effect of a decision on the processing of subsequent evidence (50–52). Participants were asked, for instance, to make a preliminary decision about the average value of a sequence of numbers. After this initial decision, they were shown more evidence before being asked to make a final decision. It was found that participants gave less weight to evidence that followed their preliminary decision, especially if that evidence was inconsistent with their preliminary decision. This pattern of findings thus resembles the confirmation bias found in more-complex decision-making contexts: Once a belief is formed, belief-inconsistent evidence is often downplayed or ignored, while consistent evidence is considered more credible (19). One way to think of this reduced sensitivity to postdecision evidence in terms of our model is as the degree to which perceived evidence affects a belief, which corresponds to the slope of the sigmoidal curves in Box 1, panel *C*. Note that the slope is high around zero, that is, where the belief is absent, indicating that, when no belief is formed yet, effects of incoming evidence are strongest. By contrast, at the belief and disbelief attractors, the slope is relatively flat, meaning that, once one of those attractors is reached, novel evidence has little effect.

Compared to simple image recognition, the waxing and waning of long-term beliefs involves a fundamentally different kind of mechanism loosely referred to as "learning" and "unlearning." In his classical 1949 book, Donald Hebb (53) suggested that, as learning takes place, the repeated simultaneous activation of weakly connected neurons gradually strengthens the synaptic connections between those neurons. So, during a phase in which new concepts are learned, the connectivity among activated neurons incrementally increases until the memory patterns corresponding to the learned concepts form a lasting trace in the brain and become attractors in the dynamics of the overall neural network. Over time, these learning processes form what we may think of as a stability landscape, with multiple attractors that each represent a different concept. All stimuli (e.g., sparrow, blackbird) associated with a given concept (e.g., bird) lead the global state of the network to flow to the same fixed attractor state.

Attractor dynamics based on Hebbian modification of network connections have become a central feature of neuronal models of memory, involving two fundamental processes: reverberations and pattern completion. Reverberation is intuitively straightforward. When you hear a word (e.g., "lemon"), the concept stays activated in your brain even if the sound of the word is already gone. Such reverberations rely on feedback in the web of recurrent excitatory connections within the pool of neurons that represents a memory. Feedback loops of the type A-triggers-B-triggers-C-triggers-A allow a brief stimulus to elicit self-sustained activity (i.e., reverberation), thus keeping the triggered concept available for some time in "working memory" (35, 43–45)—think of a good wineglass that keeps sounding for a little while after you hit it softly. Pattern completion refers to the finding that activation of a subset of the neurons involved in a memory trace can still lead to the full activation of a larger assembly, resulting in retrieval of the complete memory. In terms of our model, if a given cue is sufficiently similar to a stored memory pattern (i.e., attractor state), network activity is gradually driven to this pattern until the network fully settles on its attractor state, reflecting the retrieval of the corresponding concept (54).

The web of concepts and their properties that, together, form a person's knowledge of the world is known as "semantic memory." As a concept is activated (or "primed"), further activation spreads to neighboring concepts, increasing the probability that the associated concepts will be retrieved. In the case of the beach walk, the mere presence of the sea would have primed you to see a bird, rock, or fish, rather than a handbag. The spread of priming through a network of related concepts (sour > lemon > healthy > pills > doctor, etc.), with partially overlapping neural representations, may be modeled as a contagious spread of activation in a network of correlated attractors

(55). The effect of semantic priming is measurable when processing of a target stimulus is speeded (as measured in reaction time) by the previous presentation of a semantically related stimulus. This facilitation is important not only because semantic priming of certain concepts makes them easier to process (i.e., more "fluent") but also because fluent concepts are generally liked more and believed to be truer (56). It is not hard to see how these processes can sustain and self-reinforce beliefs and corresponding affective judgments.

Obviously, cognition cannot be fully grasped by thinking in terms of simple stability landscapes. Also, the neural realization of this system is not yet fully understood. However, it seems that beliefs are no exception to the notion that repeated activation of concepts leaves long-term memory traces, encoded through strengthening synapses, forming a deepening of a network of attractors that influences our worldview on all levels. A corollary is that erasing unwanted beliefs will typically be a slow process of unlearning, requiring repeated exposure to evidence that erodes the resilience of the belief. This is perhaps most studied in the field of psychiatry, since, as we will see, beliefs are at the basis of some of the most damaging psychiatric disorders.

How Beliefs Can Be Harmful

Although, in most societies, it is generally agreed that people should be free to believe what they want, there are, undoubtedly, beliefs that are damaging individuals or societies. Here we ask to what extent it is possible to objectively classify a belief as "harmful" and, if so, what are the prospects of getting rid of such beliefs or preventing them from taking hold in the first place. We focus on two classes of examples: pathological beliefs that play a role in mental disorders and socially embedded beliefs fueling discrimination, vaccine refusal, distortions of the political process, and violence.

Pathological Beliefs behind Psychiatric Disorders. Many mental disorders are characterized in terms of beliefs that are not connected to reality in the appropriate way (57). Often, these beliefs are resilient in the sense that they persist in the face of counterevidence and attempts at persuasion [note that this use of the term "resilience" is unrelated to common uses of the term in the clinical literature to indicate a positive state of robust mental health that persists in the face of adversity (58)]. In accordance, beliefs commonly feature as diagnostic criteria or as "symptoms" in diagnostic systems [e.g., see *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (59)]. A wide variety of beliefs have been associated with mental disorders; examples include delusional beliefs in psychosis (e.g., the belief that others can hear one's thoughts) and self-evaluation beliefs in major depression disorder (e.g., the belief that one is worthless) (60, 61).

It is noteworthy that, in many cases, the persistent nature of the belief is an important element of the diagnostic criteria specified in manuals such as the DSM-5. This is evident from the fact that temporal qualifiers are required for the relevant beliefs to be considered symptoms of mental disorders in the first place; that is, typically, they have to persist over longer periods of time (e.g., from a period of minimally 2 wk for a major depressive episode to consistent presence over adulthood for personality disorders). Hence, diagnostic systems explicitly characterize belief states as stable states, rather than as fleeting, momentary states that are occasionally present. In addition, many diagnoses require beliefs to be, in some sense, out of proportion with respect to the actual evidence. For example, the belief that one is being followed by government agencies may count as a delusional belief in psychosis, but not if one is in a situation where there is actually evidence that such risk exists (e.g., if one is a professional spy). In the latter case, the belief is not out of proportion with the evidence, and therefore may be considered normal. Thus, beliefs that are characteristically featured as symptoms of disorders are explicitly required to be both stable and out of proportion with actual evidence.

In line with our theory, pathological beliefs often interact with perceived evidence, such that the belief facilitates perception of supporting evidence. Patients suffering from anorexia nervosa, who are known to believe they are fat, have been found to misperceive their body as being fatter than it, in fact, is (62). This may, in turn, support their original belief, closing the belief trap that behaviorally promotes excessive dieting and ensuing weight loss. Patients with arachnophobia, who believe that exposure to spiders may cause harm, have been shown to overestimate the size of spiders, which can, in turn, be perceived as evidence for their belief (63). In some cases, interactions between beliefs and perceived evidence are, in fact, suspected to form the core of the disorder; a leading theory of panic attacks holds that these emerge from a feedback loop involving the perception of internal bodily states [increased heart rate, palpitation, etc. (64)] that functions as evidence for a belief (e.g., "I must be having a heart attack"). This belief promotes a fear response (flight-or-fight) that infelicitously involves the very same processes (i.e., the fear response itself promotes increased heart rate, palpitation, etc.), the result of which is a runaway feedback process that culminates in a full-blown state of panic. Clearly, fear-related beliefs may often make sense and even have an obvious evolutionary basis, as in a fear of spiders (61). They become pathological only as self-reinforcing feedbacks cause them to spiral out of proportion [e.g., arachnophobia (65)].

It is obvious, from these examples, that pathological beliefs should not be seen simply as unidirectional drivers of disorders. Rather, they are part of self-reinforcing feedbacks in pathologies. Perhaps the clearest examples of how belief trap mechanisms can stabilize disorders are found in delusions. In fact, these are almost literally characterized as belief traps, namely, as "fixed beliefs that are not amenable to change in light of conflicting evidence" (59). Stress and anxiety may stimulate the formation of delusional beliefs. For instance, an undefined anxious feeling ("delusional mood") may crystallize subsequently into a specific delusional belief (60) which may help to reduce anxiety (66). Subsequently, belief-colored perceptions may reinforce such beliefs. Whether or not a belief is really delusional is often hard to judge, but, even if it is delusional beyond doubt, a belief may spread to close relatives or friends, a phenomenon known as "induced delusional disorder," "shared psychotic disorder," or "folie-a-deux" (67). The mechanisms that drive such contagion seem unlikely to be fundamentally different from those that drive the socially held beliefs we discuss in the next section.

The strong motivation for eliminating pathological beliefs has led to systematic work exploring a wide range of options. One of the primary successful treatments of mental disorders involves the manipulation of beliefs, namely, cognitive behavioral therapy (68, 69). Techniques such as cognitive restructuring are explicitly aimed at changing beliefs (70) that serve to trigger and maintain psychopathology, which has been shown to be an effective treatment for various disorders (12). Other effective interventions (71) aimed at changing mental states, such as fear memories, involve techniques based on the idea that "after reactivation consolidated fear memories may return

to a transient labile state, requiring a process of restabilization in order to persist, [offering] a window of opportunity for modifying fear memories with amnestic agents" (72)—an almost literal description of the change processes described in the current paper. Finally, methods that utilize alternate bilateral sensory stimulation such as Eye Movement Desensitization and Reprocessing (EMDR) may be speculated to target neural mechanisms that act to destabilize associations (73), although the effectiveness and generality of these approaches in therapy is still subject to research (74).

Socially Embedded Beliefs. While the "harmfulness" of pathological beliefs is relatively uncontended, there is, by definition, no consensus about the need to eliminate socially held beliefs. Nonetheless, there is broad support for doing something about discrimination toward groups such as people of color or women. Such discrimination is often institutionalized, but it also tends to have deep roots in stereotypes held by most people that unconsciously affect their behavior (75). While the phenomenon of stereotyping is well recognized, it is notoriously difficult to do something about the negative consequences. As an arbitrary example, consider the problem of underrepresentation of women in science (76–80). Explanations include discrimination in hiring, invites for keynote lectures, peer review, and the like, but also self-selection mechanisms driving women not to choose a career in science. All of these may be largely driven by widespread implicit stereotyping.

Rational override is one prime strategy to address this problem. Well-designed protocols and close monitoring for unbiased hiring, promoting, inviting, and reviewing may eventually produce a more balanced representation. Eventually, this may neutralize stereotyping through sheer "counterevidence," a process which may be speeded up by making inspiring female role models more visible. In practice, however, inertia is notoriously large, and progress is slow. Of course, this case is just an example. We may replace "gender" with "ethnicity" or another social category and replace "science" with "politics" or another professional group and come to roughly the same conclusions. Despite many well-intended initiatives, socially embedded implicit beliefs turn out to be hard to eliminate. There is little evidence, for instance, that interventions such as workplace diversity training and media campaigns work (81). Nonetheless, some progress can be made, especially if the effect of group identity is taken into account. For instance, people are more likely to revise their attitude on gay marriage if they learn that a high-status individual in the group with whom they identify approves it (82).

A particular class of socially embedded "harmful beliefs" are conspiracy theories. Such beliefs can play an important and explicit role in the social identity of groups. Conspiracy theories generally depict established institutions or groups as hiding the truth and sustaining an unfair situation (83). This may be beneficial to some extent, as it can stimulate questioning the actions of powerful groups and promote transparency of governments (8). However, conspiracy theories may often be far from harmless, as they can be associated with political apathy, climate denial, vaccine refusal, violence, and reluctance to adhere to COVID-19 recommendations (84). Consequently, there has been much recent interest in the question of what may be done to reduce conspiracy thinking (8, 84). It is well established that ideologically homogeneous social media echo chambers can boost belief in conspiracy theories, while confronting believers with counterevidence may have weakening effects (85). This offers some hope that new forms of fact-checking and reliability labeling might, in principle, help reduce the spread of misinformation and conspiracy theories (86–88).

Nonetheless, dealing with socially held beliefs remains inherently delicate. For instance, just as in pathological beliefs, there is a gradient from normal views to unfounded conspiracy thinking. When should a critical view be labeled as a conspiracy theory? The belief that the authorities may be trusted is essential for the social contract that makes societies work (89, 90). Conspiracy theories may undermine such trust and destabilize societies (8, 84, 91). But who decides whether a socially held belief is dangerous and should be destabilized? There are plenty of historical examples of efforts to enhance stability of nations by "normalizing" cultural groups that used to be held together by shared beliefs. What may have seemed right to authorities at the time may often be judged as evil later. As an example, take the attempts to assimilate native Canadian children by moving them to residential schools. It illustrates how a whole culture may collapse when colonists insist on the cessation of cultural rituals (92). In fact, the functioning of cultural groups may fall apart upon much subtler perturbations than this. A classic example is the demise of Australian Aboriginal societies upon the introduction of stone axes given by missionaries to the converted (93). By replacing the steel axe, this practice undermined the elaborate structure of rituals and beliefs in which the old tool played a central role. Within one generation, it resulted in a complete cultural disintegration and demoralization of the group. If anything, these examples illustrate how a culture, including its beliefs, is a complex system (94) where interfering with one element can have a surprisingly disruptive impact on the whole system. While this destabilization might seem desirable in some cases, it can also have hidden perils. Despite the apparent dysfunctionality of "harmful" beliefs, they may be linked to a "deep story" which is a source of resilience for individuals and societies (95). Indeed, belief systems can make meaning of the most adverse circumstances, provide a sense of coherence, and facilitate the appraisal of new threats and opportunities (96, 97). As, in the process of changing a "harmful" element, other associated "good" elements may disintegrate, a thorough understanding of the "harmful beliefs" and their relation to broader worldviews and a degree of humility is essential as we embark on change efforts of this nature.

Outlook: Addressing Rigid Beliefs as a Social Problem

Clearly, distinguishing harmful beliefs from benign ones is far from straightforward. Nonetheless, there are some beliefs that we can safely consider unwanted. In this outlook, we first summarize what can be done to eliminate unequivocally harmful beliefs. Subsequently, we ask what societies may do to reduce the more generic problem of rigid adherence to beliefs.

Targeting Specific Harmful Beliefs. Across the examples we gave, sustained exposure to counterevidence turns out to be a fundamental component of strategies for eliminating harmful beliefs. Think of cognitive restructuring of negative beliefs about the self in psychotherapy, systematical debunking of conspiracy theories, and increasing the abundance and visibility of female role models in science. The need for sustained rather than one-time exposure is consistent with neurobiological insights in the coding of memories (98). Seeing beliefs as shaped by long-term memories implies that "unlearning" requires time. As our examples illustrate, typically, a long phase of rationally organized override is required to allow sufficient counterevidence to mount.

This is true, for instance, for cognitive behavioral training in phobic patients, but also for socially embedded harmful beliefs. For most social prejudices, sustained exposure to counterevidence basically requires societal mixing such that membership in one group (e.g., ethnicity or gender) does not predict membership in other groups (e.g., social class, political party, or profession) (99, 100). Unfortunately, the implicit prejudices are, themselves, among the forces that prevent mixing. Thus, there can be a self-amplifying feedback that persists until rationally planned institutional efforts eventually create sufficient mixing for prejudices to diminish. Polarized societies tend to move in the opposite direction, causing a societal cleavage, with attitudes on many issues becoming aligned as part of social identity (100). This makes them hard to change, because the strength of an attitude depends on its perceived importance for an individual, including its relevance for social identification (29). The link between perceived importance and the strength of a belief does imply, however, that approaches aimed at playing down the importance of a harmful belief and ameliorating the polarization around it might help reduce its prevalence in the face of counterevidence (28, 101).

A More Generic Approach: Reducing Dichotomous Thinking. While "curing" specific beliefs that are unequivocally harmful makes sense, it is also worth asking how the emergence of rigid beliefs, in general, may be prevented. After all, irrespective of the specific content of beliefs, excessive rigidity is, in many ways, an impediment to human progress, as pointed out by philosophers and scientists for centuries (3, 4). In our view, probably the most important element to consider in this generic context is the tendency for dichotomous thinking, also referred to as black-and-white thinking, binary thinking, or absolutist thinking. It is the tendency to think in terms of binary oppositions (i.e., "black or white," "good or bad," "all or nothing"). The tendency for dichotomous thinking may be assessed using an inventory where participants mark how strongly they agree with a list of statements (48, 102). Alternatively, a tendency for dichotomization may be inferred from the excessive use of absolute words in natural language (e.g., always, never, totally, ever, never, and must) (103, 104). The much-studied character trait of intolerance to ambiguity correlates with such dichotomous thinking (105).

Our model presents dichotomization as an essential driver of belief resilience (Box 1, panel *A*). Indeed, a central role of black-and-white thinking is consistent with the observation that dichotomization is associated with belief-related problems ranging from social prejudices (106, 107) to psychiatric disorders (48, 49). It is also in line with the finding that tolerance to ambiguity correlates with intellectual humility, the capacity to recognize one's fallibility and to update beliefs in the face of new evidence (13). In psychology, a well-known technique for stimulating a person to explore and resolve ambiguity is motivational interviewing (108). This approach involves 1) empathizing with, as opposed to challenging, the person being interviewed; 2) surfacing any discrepancy between the subject's current and desired behavior; 3) encouraging interviewees to explore and expand on their feelings, particularly those that are dissonant; and 4) supporting self-efficacy, that is, the confidence in their ability to change. While this technique is mostly used to help clients in psychological practice, it has also been used successfully in maternity wards, to reduce vaccine hesitancy (109). In cognitive behavioral therapy (12, 110), a related technique to invite ambiguity is having the patient come up with alternative views. For instance, one intervention in the treatment of panic is to teach the patient to come up with at least one alternative explanation for the somatic arousal in addition to the problematic one ("I am having a heart attack"), for instance, "I am nervous," or "I was climbing stairs."

Important clues for more-generic ways to reduce black-and-white thinking in societies come from studies (48, 49) suggesting that black-and-white thinking happens more when there is a lack of cognitive resources, a condition which is associated with stress (111) as well as poor educational background (112). The link to education could have various causal explanations, but one obvious possibility is that education provides skills for the use of a rational approach, the essential machinery for overriding rigidly held beliefs. Of course, it may also help to provide balanced views early on. It has been shown that early "inoculation" with correct information may help to prevent conspiracy theories from taking hold (8). This is consistent with the well-documented primacy effect in confirmation bias studies (19). Framing it somewhat cynically, seeding "good beliefs" early helps prevent harmful beliefs from taking hold later.

The important finding that stress promotes black-and-white thinking resonates with the neurobiological finding that inhibition-driven competition between populations of neurons representing alternative interpretations is amplified during states of increased arousal or stress (46, 47). Stress-boosted dichotomous thinking may thus well be a fundamental driver of belief rigidity. This helps to clarify the importance of empathetic listening and relationships, but the dominant role of stress also fits the observation that conspiracy theories tend to originate in times of uncertainty and crisis (8, 113), and that the same is true for mental disorders (114, 115).

A Social Approach to Harmful Beliefs. The insights into the drivers of dichotomous thinking suggest an important role for societal processes in driving belief inertia (Fig. 4).

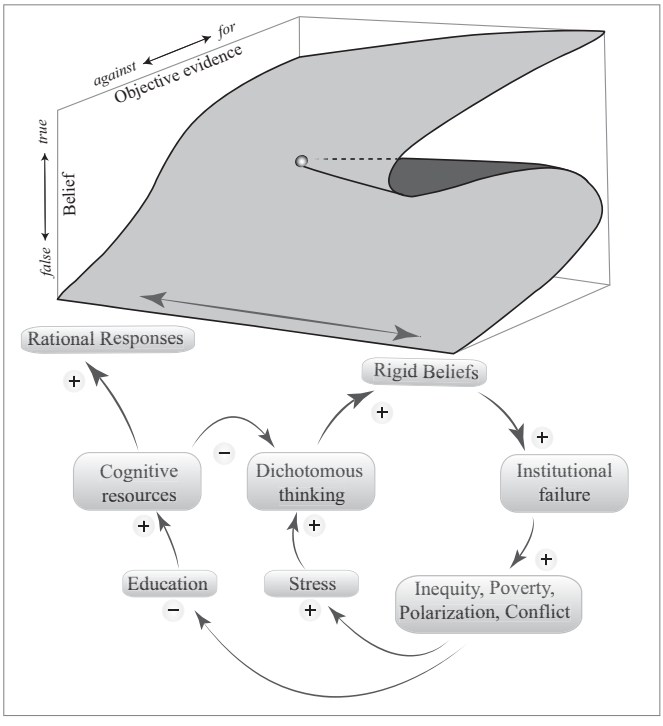


Fig. 4. Mechanisms through which rigidity of beliefs and societal failure may reinforce each other. The documented roles of cognitive resources and dichotomous thinking imply that mitigation of rigid harmful beliefs may require improving the educational system and addressing problems of inequity, poverty, polarization, and conflict.

Dichotomous thinking will be more prevalent in societies where people are stressed and poorly educated. A resulting rise in pathological beliefs, conspiracy thinking, and social prejudices may, in turn, hamper societal thriving, thus implying the potential for a self-reinforcing feedback toward societal failure. While targeting specific harmful beliefs with counterevidence is important, the broader view thus implies that a generic societal approach may help address belief resilience at the roots. The evidence we presented suggests that the prevalence of rigid beliefs may perhaps best be mitigated by strengthening educational systems and addressing inequity and the related problems of poverty, conflict, food insecurity, and social cleavage. Put bluntly, measures such as a universal base income might go a surprisingly long way in reducing the resilience of harmful beliefs. This would be consistent with the finding that interventions similar to basic income tend to reduce mental health problems (116), and resonates with a two-component model of belief in conspiracy theories where inequities, prejudices, and breaches of trust in authorities send individuals searching "down the rabbit hole" where misinformation can become reinforced by confirmation bias and the "posttruth" dynamics of echo chambers (117).

In conclusion, while beliefs are a necessary element of healthy cognition, rigid beliefs are the basis of some of the most damaging problems faced by humans both on an individual and on a social level. A surprisingly universal image of what shapes belief resilience emerges from elements provided by disciplines ranging from neurobiology to sociology. Two elements are central when it comes to the question of what we can do about harmful beliefs:

1) Reducing the resilience of specific harmful beliefs requires sustained exposure to counterevidence, which typically requires organized rational override.
2) Reducing rigidity of beliefs in general can be achieved through improving education and reducing social stress.

Meanwhile, it is worth considering the possibility that a socially held attitude toward the general relevance of evidence may have an effect on belief resilience. Indeed, the strength of the "metabelief" that beliefs should change according to evidence appears predictive for peoples' attitude toward conspiracy theories; moral, political, and religious perspectives; and faith in science (118). The prevalence of this metabelief may change over time. Indeed, the surge of fact-free, posttruth political argumentation suggests that the balance between the roles of emotion and reasoning is changing (119–121).

In view of the complementary contributions from neurobiology, psychology, and social sciences, cross-disciplinary efforts may be an exciting way forward in finding ways to overcome the prevalence of resilient harmful beliefs and the many problems they cause.

Data Availability. All study data are included in the main text.

ACKNOWLEDGMENTS. We thank Naomi Ellemers, John Robinson, Kenneth Kendler, and Leor Zmigrod for insightful comments that helped us sharpen the manuscript.

1. E. Schwitzgebel, Belief. The Stanford Encyclopedia of Philosophy (Winter 2019). https://plato.stanford.edu/archives/win2019/entries/belief/. Accessed 1 July 2022.
2. J. Leicester, The nature and purpose of belief. *J. Mind Behav.* **29**, 217–237 (2008).
3. B. Spinoza, *The Essential Spinoza: Ethics and Related Writings* (Hackett, 2006).
4. S. Pinker, *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress* (Penguin, 2018).
5. E. West, M. Krychman, Natural aphrodisiacs–A review of selected sexual enhancers. *Sex. Med. Rev.* **3**, 279–288 (2015).
6. A. E. Cruz-Inigo, B. Ladizinski, A. Sethi, Albinism in Africa: Stigma, slaughter and awareness campaigns. *Dermatol. Clin.* **29**, 79–87 (2011).
7. J. Goodare, *The European Witch-Hunt* (Routledge, 2016).
8. K. M. Douglas *et al.*, Understanding conspiracy theories. *Polit. Psychol.* **40**, 3–35 (2019).
9. E. L. Olson, The rationalization and persistence of organic food beliefs in the face of contrary evidence. *J. Clean. Prod.* **140**, 1007–1013 (2017).
10. J. Haidt, The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* **108**, 814–834 (2001).
11. D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011).
12. S. G. Hofmann, A. Asnaani, I. J. Vonk, A. T. Sawyer, A. Fang, The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognit. Ther. Res.* **36**, 427–440 (2012).
13. L. Zmigrod, S. Zmigrod, P. J. Rentfrow, T. W. Robbins, The psychological roots of intellectual humility: The role of intelligence and cognitive flexibility. *Pers. Individ. Dif.* **141**, 200–208 (2019).
14. M. Scheffer, F. R. Westley, The evolutionary basis of rigidity: Locks in cells, minds, and society. *Ecol. Soc.* **12**, 36 (2007).
15. K. Nyborg *et al.*, Social norms as solutions. *Science* **354**, 42–43 (2016).
16. E. Bakshy, S. Messing, L. A. Adamic, Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
17. I. C. McManus, M. Freegard, J. Moore, R. Rawles, Science in the Making: Right Hand, Left Hand. II: The duck-rabbit figure. *Laterality* **15**, 166–185 (2010).
18. M. Scheffer, *Critical Transitions in Nature and Society* (Princeton University Press, Princeton, NJ, 2009).
19. R. S. Nickerson, Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **2**, 175 (1998).
20. F. Bacon, *Francis Bacon: The New Organon* (Cambridge University Press, 2000).
21. P. Thagard, "Mapping minds across cultures" in *Grounding Social Sciences in Cognitive Sciences*, R. Sun, Ed. (MIT Press, Cambridge, MA, 2011), pp. 35–62.
22. P. Thagard, K. Verbeurgt, Coherence as constraint satisfaction. *Cogn. Sci.* **22**, 1–24 (1998).
23. L. Festinger, *A Theory of Cognitive Dissonance* (Stanford University Press, 1962).
24. M. Feinberg, R. Willer, Apocalypse soon? Dire messages reduce belief in global warming by contradicting just-world beliefs. *Psychol. Sci.* **22**, 34–38 (2011).
25. T. Homer-Dixon *et al.*, A complex systems approach to the study of ideology: Cognitive-affective structures and the dynamics of belief systems. *J. Soc. Polit. Psych.* **1**, 337–363 (2013).
26. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
27. M. Scheffer, F. Westley, W. Brock, Slow response of societies to new problems: Causes and costs. *Ecosystems (N. Y.)* **6**, 493–502 (2003).
28. H. L. van der Maas, J. Dalege, L. Waldorp, The polarization within and across individuals: The hierarchical Ising opinion model. *J. Complex Netw.* **8**, cnaa010 (2020).
29. L. C. Howe, J. A. Krosnick, Attitude strength. *Annu. Rev. Psychol.* **68**, 327–351 (2017).
30. L. Festinger, H. W. Riecken, S. Schachter, *When Prophecy Fails* (Harper and Row, New York, 1956).
31. E. E. Harmon-Jones, *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology* (American Psychological Association, 2019).
32. S. A. Thompson, Three weeks inside a pro-Trump QANON chat room. *New York Times*, 26 January 2021. https://www.nytimes.com/interactive/2021/01/26/opinion/trump-qanon-washington-capitol-hill.html. Accessed 1 July 2022.
33. D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, 1992).
34. M. P. Karlsson, D. G. Tervo, A. Y. Karpova, Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* **338**, 135–139 (2012).
35. X.-J. Wang, Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).
36. C. M. Warren, S. Nieuwenhuis, T. H. Donner, Perceptual choice boosts network stability: Effect of neuromodulation? *Trends Cogn. Sci.* **19**, 362–364 (2015).
37. A. Schurger, I. Sarigiannidis, L. Naccache, J. D. Sitt, S. Dehaene, Cortical activity is more stable when sensory stimuli are consciously perceived. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2083–E2092 (2015).
38. D. J. Amit, The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behav. Brain Sci.* **18**, 617–625 (1995).
39. D. Durstewitz, N. M. Vittoz, S. B. Floresco, J. K. Seamans, Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* **66**, 438–448 (2010).
40. A. Finkelstein *et al.*, Attractor dynamics gate cortical information flow during decision-making. *Nat. Neurosci.* **24**, 843–850 (2021).
41. D. Peixoto *et al.*, Decoding and perturbing decision states in real time. *Nature* **591**, 604–609 (2021).
42. M. Scheffer *et al.*, Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
43. E. Düzel, W. D. Penny, N. Burgess, Brain oscillations and memory. *Curr. Opin. Neurobiol.* **20**, 143–149 (2010).
44. T. J. Wills, C. Lever, F. Cacucci, N. Burgess, J. O'Keefe, Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873–876 (2005).
45. G. Deco, E. T. Rolls, R. Romo, Stochastic dynamics as a principle of brain function. *Prog. Neurobiol.* **88**, 1–16 (2009).
46. E. Eldar, J. D. Cohen, Y. Niv, The effects of neural gain on attention and learning. *Nat. Neurosci.* **16**, 1146–1153 (2013).
47. M. Mather, M. R. Sutherland, Arousal-biased competition in perception and memory. *Perspect. Psychol. Sci.* **6**, 114–133 (2011).
48. T. Mieda, K. Taku, A. Oshio, Dichotomous thinking and cognitive ability. *Pers. Individ. Dif.* **169**, 110008 (2021).
49. A. Oshio, T. Mieda, K. Taku, Younger people, and stronger effects of all-or-nothing thoughts on aggression: Moderating effects of age on the relationships between dichotomous thinking and aggression. *Cogent Psychol.* **3**, 1244874 (2016).
50. B. C. Talluri, A. E. Urai, K. Tsetsos, M. Usher, T. H. Donner, Confirmation bias through selective overweighting of choice-consistent evidence. *Curr. Biol.* **28**, 3128–3135.e8 (2018).
51. Z. Z. Bronfman *et al.*, Decisions reduce sensitivity to subsequent information. *Proc. R. Soc. B Biol. Sci.* **282**, 20150228 (2015).
52. M. Rollwage *et al.*, Confidence drives a neural confirmation bias. *Nat. Commun.* **11**, 2634 (2020).
53. D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Psychology, 2005).
54. V. Daelli, A. Treves, Neural attractor dynamics in object recognition. *Exp. Brain Res.* **203**, 241–248 (2010).
55. I. Lerner, S. Bentin, O. Shriki, Spreading activation in an attractor network with latching dynamics: Automatic semantic priming revisited. *Cogn. Sci.* **36**, 1339–1382 (2012).

56. A. L. Alter, D. M. Oppenheimer, Uniting the tribes of fluency to form a metacognitive nation. *Pers. Soc. Psychol. Rev.* **13**, 219–235 (2009).
57. D. Borsboom, A. O. J. Cramer, A. Kalis, Reductionism in retreat. *Behav. Brain Sci.* **42**, e32 (2019).
58. R. Kalisch *et al.*, Deconstructing and reconstructing resilience: A dynamic network approach. *Perspect. Psychol. Sci.* **14**, 765–777 (2019).
59. A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (American Psychiatric, 2013).
60. M. G. Henriksen, J. Parnas, "Delusional mood" *The Oxford Handbook of Phenomenological Psychopathology*, G. Stanghellini *et al.*, Eds. (Oxford University Press, 2019). pp. 465–474.
61. A. V. Horwitz, J. C. Wakefield, *All We Have to Fear: Psychiatry's Transformation of Natural Anxieties into Mental Disorders* (Oxford University Press, 2012).
62. K. L. Challinor *et al.*, Body size and shape misperception and visual adaptation: An overview of an emerging research paradigm. *J. Int. Med. Res.* **45**, 2001–2008 (2017).
63. T. Leibovich, N. Cohen, A. Henik, Itsy bitsy spider?: Valence and self-relevance predict size estimation. *Biol. Psychol.* **121**, 138–145 (2016).
64. M. G. Craske *et al.*; DSM V Anxiety; OC Spectrum; Posttraumatic and Dissociative Disorder Work Group, Panic disorder: A review of DSM-IV panic disorder and proposals for DSM-V. *Depress. Anxiety* **27**, 93–112 (2010).
65. S. J. Thorpe, P. M. Salkovskis, Phobic beliefs: Do cognitive factors play a role in specific phobias? *Behav. Res. Ther.* **33**, 805–816 (1995).
66. D. Freeman, P. A. Garety, Worry, worry processes and dimensions of delusions: An exploratory investigation of a role for anxiety processes in the maintenance of delusional distress. *Behav. Cogn. Psychother.* **27**, 47–62 (1999).
67. D. Arnone, A. Patel, G. M.-Y. Tan, The nosological significance of folie à deux: A review of the literature. *Ann. Gen. Psychiatry* **5**, 11 (2006).
68. A. T. Beck, Cognitive therapy: Nature and relation to behavior therapy. *Behav. Ther.* **1**, 184–200 (1970).
69. A. Ellis, *Reason and Emotion in Psychotherapy* (Citadel, 1962).
70. A. T. Beck, E. A. Haigh, Advances in cognitive theory and therapy: The generic cognitive model. *Annu. Rev. Clin. Psychol.* **10**, 1–24 (2014).
71. M. Soeter, M. Kindt, An abrupt transformation of phobic behavior after a post-retrieval amnesic agent. *Biol. Psychiatry* **78**, 880–886 (2015).
72. M. Kindt, The surprising subtleties of changing fear memory: A challenge for translational science. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170033 (2018).
73. J. Baek *et al.*, Neural circuits underlying a psychotherapeutic regimen for fear disorders. *Nature* **566**, 339–343 (2019).
74. P. Cuijpers, S. C. V. Veen, M. Sijbrandij, W. Yoder, I. A. Cristea, Eye movement desensitization and reprocessing for mental health problems: A systematic review and meta-analysis. *Cogn. Behav. Ther.* **49**, 165–180 (2020).
75. D. J. Schneider, *The Psychology of Stereotyping* (Guilford, 2005).
76. K. R. O'Brien, M. Scheffer, E. H. van Nes, R. van der Lee, How to break the cycle of low workforce diversity: A model for change. *PLoS One* **10**, e0133208 (2015).
77. H. Shen, Inequality quantified: Mind the gender gap. *Nature* **495**, 22–24 (2013).
78. V. Larivière, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: Global gender disparities in science. *Nature* **504**, 211–213 (2013).
79. S. J. Ceci, W. M. Williams, Understanding current causes of women's underrepresentation in science. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3157–3162 (2011).
80. R. van der Lee, N. Ellemers, Gender contributes to personal research funding success in The Netherlands. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12349–12353 (2015).
81. E. L. Paluck, D. P. Green, Prejudice reduction: What works? A review and assessment of research and practice. *Annu. Rev. Psychol.* **60**, 339–367 (2009).
82. B. F. Harrison, M. R. Michelson, *Listen, We Need to Talk: How to Change Attitudes about LGBT Rights* (Oxford University Press, 2017).
83. J.-W. Van Prooijen, A. P. Krouwel, T. V. Pollet, Political extremism predicts belief in conspiracy theories. *Soc. Psychol. Personal. Sci.* **6**, 570–578 (2015).
84. K. M. Douglas, Are conspiracy theories harmless? *Span. J. Psychol.* **24**, e13 (2021).
85. B. R. Warner, R. Neville-Shepard, Echoes of a conspiracy: Birthers, truthers, and the cultivation of extremism. *Commun. Q.* **62**, 1–17 (2014).
86. N. Hassan, M. Yousuf, M. Mahfuzul Haque, J. A. Suarez Rivas, M. K. Islam, "Examining the roles of automation, crowds and professionals towards sustainable fact-checking" in *Companion Proceedings of The 2019 World Wide Web Conference* L. Liu, R. White, Eds (Association for Computing Machinery, New York, NY, 2019), pp. 1001–1006.
87. G. Pennycook, D. G. Rand, Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2521–2526 (2019).
88. N. Walter, J. Cohen, R. L. Holbert, Y. Morag, Fact-checking: A meta-analysis of what works and for whom. *Polit. Commun.* **37**, 350–375 (2019).
89. T. Besley, State capacity, reciprocity, and the social contract. *Econometrica* **88**, 1307–1335 (2020).
90. L. Korn, R. Böhm, N. W. Meier, C. Betsch, Vaccination as a social contract. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14890–14899 (2020).
91. B. Albertson, K. Guiler, Conspiracy theories, election rigging, and support for democratic norms. *Res. Politics* **7**, https://doi.org/10.1177/2053168020959859 (2020).
92. D. B. MacDonald, G. Hudson, The genocide question and Indian residential schools in Canada. *Can. J. Political Sci.* **45**, 427–449 (2012).
93. L. Sharp, Steel axes for stone-age Australians. *Hum. Organ.* **11**, 17–22 (1952).
94. C. Geertz, *Local Knowledge: Further Essays in Interpretive Anthropology* (Basic, 2008).
95. A. R. Hochschild, *Strangers in Their Own Land: Anger and Mourning on the American Right* (The New Press, 2018).
96. F. Walsh, "Family resilience: A dynamic systemic framework" in *Multisystemic Resilience: Adaptation and Transformation in Contexts of Change*, M. Ungar, Ed. (Oxford University Press, 2021), pp. 255–270.
97. A. Antonovsky, *Unraveling the Mystery of Health: How People Manage Stress and Stay Well* (Jossey-Bass, 1987).
98. J. J. Langille, R. E. Brown, The synaptic theory of memory: A historical survey and reconciliation of recent opposition. *Front. Syst. Neurosci.* **12**, 52 (2018).
99. D. Baldassarri, M. Abascal, Diversity and prosocial behavior. *Science* **369**, 1183–1187 (2020).
100. E. Klein, *Why We're Polarized* (Simon and Schuster, 2020).
101. J. Dalege, D. Borsboom, F. van Harreveld, H. L. van der Maas, The attitudinal entropy (AE) framework as a general theory of individual attitudes. *Psychol. Inq.* **29**, 175–193 (2018).
102. A. Oshio, Development and validation of the dichotomous thinking inventory. *Soc. Behav. Personal.* **37**, 729–741 (2009).
103. M. Al-Mosaiwi, T. Johnstone, In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clin. Psychol. Sci.* **6**, 529–542 (2018).
104. J. Bollen *et al.*, Historical language records reveal a surge of cognitive distortions in recent decades. *Proc. Natl. Acad. Sci. U.S.A.* **30**, 118 (2021).
105. M. Lauriola, R. Foschi, O. Mosca, J. Weller, Attitude toward ambiguity: Empirically robust factors in self-report personality scales. *Assessment* **23**, 353–373 (2016).
106. E. Forsberg, A. Nilsson, Ø. Jørgensen, Moral dichotomization at the heart of prejudice: The role of moral foundations and intolerance of ambiguity in generalized prejudice. *Soc. Psychol. Personal. Sci.* **10**, 1002–1010 (2019).
107. A. Roets, A. Van Hiel, Allport's prejudiced personality today: Need for closure as the motivated cognitive basis of prejudice. *Curr. Dir. Psychol. Sci.* **20**, 349–354 (2011).
108. W. R. Miller, S. Rollnick, *Motivational Interviewing: Helping People Change* (Guilford, 2012).
109. A. Gagneur *et al.*, A postpartum vaccination promotion intervention using motivational interviewing techniques improves short-term vaccine coverage: PromoVac study. *BMC Public Health* **18**, 811 (2018).
110. A. T. Beck, D. D. Davis, A. Freeman, *Cognitive Therapy of Personality Disorders* (Guilford, 2015).
111. A. Boals, J. B. Banks, Effects of traumatic stress and perceived stress on everyday cognitive functioning. *Cogn. Emotion* **26**, 1335–1343 (2012).
112. S. J. Ritchie, E. M. Tucker-Drob, How much does education improve intelligence? A meta-analysis. *Psychol. Sci.* **29**, 1358–1369 (2018).
113. J.-W. van Prooijen, K. M. Douglas, Conspiracy theories as part of history: The role of societal crisis situations. *Mem. Stud.* **10**, 323–333 (2017).
114. H. Yao, J.-H. Chen, Y.-F. Xu, Patients with mental health disorders in the COVID-19 epidemic. *Lancet Psychiatry* **7**, e21 (2020).
115. B. Pfefferbaum, C. S. North, Mental health and the Covid-19 pandemic. *N. Engl. J. Med.* **383**, 510–512 (2020).
116. M. Gibson, W. Hearty, P. Craig, The public health effects of interventions similar to basic income: A scoping review. *Lancet Public Health* **5**, e165–e176 (2020).
117. J. M. Pierre, Mistrust and misinformation: A two-component, socio-epistemic model of belief in conspiracy theories. *J. Soc. Polit. Psych.* **8**, 617–641 (2020).
118. G. Pennycook, J. A. Cheyne, D. J. Koehler, J. A. Fugelsang, On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgm. Decis. Mak.* **15**, 476 (2020).
119. F. Fischer, Knowledge politics and post-truth in climate denial: On the social construction of alternative facts. *Crit. Policy Stud.* **13**, 133–152 (2019).
120. A. Kofman, Bruno Latour, the post-truth philosopher, mounts a defense of science. *The New York Times Magazine*, 25 October 2018. https://www.nytimes.com/2018/10/25/magazine/bruno-latour-post-truth-philosopher-science.html. Accessed 1 July 2022.
121. S. Lewandowsky, U. K. Ecker, J. Cook, Beyond misinformation: Understanding and coping with the "post-truth" era. *J. Appl. Res. Mem. Cogn.* **6**, 353–369 (2017).