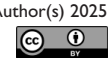
Theory & Psychology

Article

The looping effects of psychological theories: From anomaly to opportunity

Theory & Psychology
1–22


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09593543251381278
journals.sagepub.com/home/tap
Sage

Alex Gillespie
London School of Economics
Oslo New University

Brady Wagoner
University of Copenhagen
Aalborg University

Abstract

Looping effects were introduced by philosopher Ian Hacking to refer to the two-way interactive relationship between scientific classifications and the people classified. This article extends the concept beyond classification to consider how psychological theories shape and transform their phenomena through looping effects, rather than merely mirroring them. These effects have been conceptualized as destabilizing, problematic, and anomalous. In contrast, we argue that looping effects should be brought within our theorizing and harnessed. To this end, we introduce a two-by-two typology of looping effects, differentiating between weakening versus strengthening effects and intended versus unintended effects. Looping can lead some theories to be resisted and thus weakened (e.g., persuasion, nudging) and other theories to be adopted and thus strengthened (e.g., mnemonic or metacognitive strategies). While most theories begin with unintended looping effects, they often develop towards intentional looping effects (e.g., aimed at weakening cognitive biases or strengthening heuristics). We argue that looping effects are a significant yet under-studied source of variance in human behavior. Psychological theories do not merely refer to causal mechanisms; theories can also be causes of human behavior. We make practical suggestions for incorporating looping effects into psychological theory, thereby enhancing its robustness and societal relevance.

Keywords

backfire effects, boomerang effects, looping effects, psychological theory, reactance

Corresponding author:
Alex Gillespie, Department of Psychological and Behavioural Science, London School of Economics, Houghton Street, London WC2A 2AE, UK.
Email: a.t.gillespie@lse.ac.uk

Introduction

> Some gentleman with an ignoble, or rather a derisive and reactionary air, springs up suddenly out of nowhere, puts his arms akimbo and says to all of us, "Come on, gentlemen, why shouldn't we get rid of all this calm reasonableness with one kick, just so as to send all these logarithms to the devil. (Dostoevsky, 1864, p. 33)

Many psychological theories interact with the phenomenon being theorized. While it is often assumed that our theories are like impartial mirrors of nature (Gergen, 2015), history shows that our theories not only become part of common sense (Moscovici, 2008) but can also transform the underlying phenomena. People who learn about biases can avoid them (Morewedge et al., 2015), and people taught about persuasion techniques can resist them (Roozenbeek et al., 2022). Meanwhile, when people learn about decision-making heuristics, mnemonic strategies, or metacognition skills, the prevalence and efficacy of these phenomena can be strengthened (Rivas et al., 2022). Why does awareness of a theory sometimes weaken it and, at other times, strengthen it?

There is a tension in the aims of psychology. On the one hand, it aims to develop timeless causal theories that explain and predict human behavior and cognition with numeric accuracy (Meehl, 1990). On the other hand, it aims to enhance human behavior and cognition (Zimbardo, 2004). But, the timelessness of the former risks being undermined by the transformative potential of the latter. Recent concern about the lack of robust theory in psychology (Eronen & Bringmann, 2021) has failed to question how theory can be robust when it transforms the phenomena being theorized. Our commitment to timeless rule-like theories has inhibited our ability to create dynamic world-making, and human-making, theories (Power et al., 2023). Taking psychology's transformative potential seriously implies reconceptualizing what robust theory looks like.

This tension is also evident in the replication crisis, namely the finding that many psychological results, especially in social psychology, fail to replicate (Open Science Collaboration, 2015). This crisis has prompted more robust methods, pre-registering studies, making data open access, and multisite replications, albeit with mixed results (Baumeister et al., 2023). While it is necessary to improve research rigor, we should not expect all theories to replicate (Baucal et al., 2020). Indeed, it might be a triumph to attenuate undesirable patterns of behavior and cognition (e.g., biases, distressing cognitions, sources of conflict). If the aim of psychology is to contribute to society and transform human behavior, then naive replication must be reconceptualized.

We introduce a typology of *looping effects* in psychology to conceptualize how psychological theories interact with the phenomena being theorized. Specifically, we distinguish between looping that either weakens or strengthens a theory, and looping that is either unintended or intended. We argue that theories that aim to predict and control are more likely to weaken with looping (e.g., will be resisted), while theories that aim to empower are more likely to strengthen (e.g., will be used and refined). Moreover, we reveal a tendency for psychological theory, as it matures, to move from having unintended looping effects to having increasingly intentional effects (e.g., training people about the theory). By conceptualizing and mastering looping effects, psychology can produce useful theories that not only enrich human behavior but also spread and strengthen over time.

Psychology: A natural or a social science?

The idea that looping effects are problematic and even embarrassing stems from suppressing psychology's historical and cultural aspects. A dominant assumption since the Enlightenment has been that the mind is unchanging across time and space. According to this view, true knowledge about both mind and matter is derived from universally valid reasoning (e.g., Descartes, 1641; Kant, 1871). This assumption led theorists to emphasize the law-like nature of cognition, even specifying it in mathematical terms (e.g., Chomsky, 1995; Fodor, 1983).

Yet, from the outset, this assumption was criticized. Vico (1725) argued that knowledge is made, not found. Similarly, Herder (1784) grounded cognition within culture, arguing that each culture has a distinctive mentality that needs to be understood and assessed from within its own principles. This challenged his mentor Kant's notion of universal human rationality and celebrated the variability of thought across time and place.

Wundt (1912), regarded as the founder of modern psychology, tried to integrate the natural and historical sides of the discipline (Farr, 1983). He distinguished between *experimental psychology*, which studied basic psychological functions using natural science techniques, and *Völkerpsychologie*, which studied 'higher psychological functions' through a comparative-developmental analysis of communities and their culture (language, myth, customs). Wundt used the term 'higher' in two senses: first, higher functions are built on top of lower physiological processes; second, they regulate the lower processes. Crucially, for our purposes, the lower psychological functions are universal while the higher functions are situated within a particular time and place (i.e., language-mediated). It follows from Wundt's distinction that any higher mental functions that rely upon language and culture will change over time and, thus, potentially be impacted by looping effects.

The emerging discipline of psychology embraced Wundt's experimental psychology but rejected his *Völkerpsychologie* (Danziger, 1990). Instead of grappling with the complexity of studying psychological processes as historical, increasing swathes of psychology were assumed to comprise natural science phenomena. But, although this approach yielded short-term benefits (sidestepping contextual complexity), it set up a paradigm that made cultural and historical variations in psychological phenomena an anomaly.

According to Kuhn (1962), anomalies herald scientific revolutions. Anomalies are deviations from expectation that, although widely acknowledged, are neglected because they do not fit the dominant paradigm. Arguably, the characterization of all psychology as a natural science, rejecting any historical-cultural aspect, inadvertently transformed cultural-historical variation in psychological functioning into anomalies. The recent concern with Western, educated, industrialized, rich, and democratic (WEIRD) populations is belated recognition of this variability (Henrich et al., 2010). What is less often recognized and much more anomalous is the fact that psychological knowledge sometimes transforms its object.

Social psychology as history

There are two ways in which psychological theory is entangled with human culture. First, the social context shapes our theories. For example, biomedical, psychiatric, and

neuroscientific knowledge is shaped by corporate interests, often using technological advances to increase data accumulation, reduce risk, and protect patents (Rose, 2018). Second, the circulation of psychological knowledge in society has consequences. For example, ideas about psychotherapy, behavioral reinforcement, and unconscious processes have shaped how people regulate themselves (Zimbardo, 2004). While the impact of psychology in society has been broad, our focus is on a sub-set of these downstream effects: where psychological theory shapes the phenomena being theorized.

Gergen (1973) was among the first to argue that psychological theories often altered the behavior they aimed to understand and predict. He argued that science and society were in a feedback loop, with each constituting the other. Becoming predictable to others, he argued, is a vulnerability that will be resisted.

> We are satisfied with principles of attitude change until we find them being used in information campaigns dedicated to changing our behavior. At this point, we may feel resentful and react recalcitrantly. The more potent the theory is in predicting behavior, the broader its public dissemination and the more prevalent and resounding the reaction. Thus, strong theories may be subject to more rapid invalidation than weak ones. (Gergen, 1973, p. 314)

Gergen pointed to core concepts being loaded with social judgment. For example, measures of authoritarianism, closed-mindedness, Machiavellianism, and obedience all imply a desirable-to-undesirable continuum. Knowledge of these concepts and their social judgment will, at the very least, cause social desirability effects, but it might also lead to a transformation of the phenomenon itself. For example, a high score on closed-mindedness might lead someone to become less closed-minded, thus undermining the validity of measurement.

Although Gergen's article became a classic, it also attracted criticism. A review of the debate concluded that his argument was overstated and that contextualism could be brought within the positivistic paradigm: "Gergen's pessimism is no longer warranted and the revolution he proclaimed no longer needed" (Blank, 1988, p. 651). Yet, the looping dynamics that Gergen identified have become increasingly anomalous.

Contemporary psychological theory now permeates news articles, self-help books, viral videos, social media posts, documentaries, and films. All major developments in psychology, from priming techniques to cognitive-behavioral therapy, and from minimal group paradigm games to power-poses, have been widely disseminated. One could ask: Are any participants left who remain naive to basic psychological insights?

Moreover, many of the participants in psychological research are psychology students who have been trained in psychology (Arnett, 2016), creating the risk of even tighter looping, where students are being taught the theories that they are producing data on (Sadeghiyeh et al., 2021). Evidence suggests this can have an impact. For example, in experiments, economics students tend to conform more to economic theory than other students (Gerlach, 2017), indicating the strengthening of the theories through training.

Gergen's (1973) article was, we suggest, ahead of its time and broadly correct. Our contribution is to develop Gergen's (1973, p. 314) provocative claim that "strong theories may be subject to more rapid invalidation than weak ones." Specifically, we propose that the theories most likely to be unintentionally weakened by looping are those that aim

to predict and control. In contrast, useful theories that empower people are unlikely to be invalidated through public dissemination; indeed, they may even be strengthened.

Looping effects

Hacking (1995, 1999) coined the term looping effects to conceptualize how social scientific categories interact with the people and behaviors they refer to. He argued that the natural sciences study *natural kinds* that don't change depending on how they are categorized. The social sciences, however, study *human kinds* (or *interactive kinds*), which tend to change in reaction to how we categorize them. For example, popularizing the classification of a mental illness (e.g., multiple personality disorder) can lead to increased presentations and diagnoses of the illness.

Critics have challenged a clear-cut distinction between natural and human kinds on a number of grounds, including that natural kinds may also be interactive (e.g., labeling a species 'endangered' or a 'pest' can change them in the long run). However, Laimann (2020) defends the distinction by arguing that human kinds are not so much 'interactive' as 'capricious'. By this she means that they change in unexpected ways as a function of their grounding in both natural and social processes (or 'status kinds'). For our purposes, it is enough to reiterate Hacking's original argument that human kinds are reactive: the act of classification can impact the phenomenon itself.

For example, theories of mental illness, when popularized, give people a language to describe their own symptoms to doctors, thus potentially reinforcing the classification. Hacking (1995) himself powerfully detailed how multiple personality disorder dramatically increased in both frequency and severity after it became widely known through media representations of it. Looping effects in psychological diagnosis have become increasingly salient. It is now common for people to self-diagnose via social media, popular books, AI, or even scientific articles before seeking a professional diagnosis (Underhill & Foulkes, 2024). The problem is that awareness of diagnostic criteria might impact the diagnostic process itself (Brinkmann, 2016). This raises fundamental questions: How does a prior self-diagnosis interact with a professional diagnosis? And, is a diagnosis equally valid if the client is familiar or unfamiliar with the diagnostic criteria?

The looping effects of categorizations extend far beyond diagnosis. For example, Becker (1963) showed how the label of being deviant could potentially spoil people's identity. They might try to disprove the label, by avoiding deviant behaviors. Or, they might embrace the label, becoming increasingly deviant. The label can also change how other people interact with the person thus classified (Heasman & Gillespie, 2019). There is a similar debate in mental health about the secondary effects of categorizations on people's self-perception and actions (Lindholm & Wickström, 2020). The consequences of categorizations can be unexpected, and even the attempts to craft desirable categorizations can have unintended consequences (Corrigan, 2016).

Categorizations can also have inter-group looping effects. Social identity theory has long shown how categorizations can create groups and contribute to conflict (Brown, 2000). Such groupings can even be created with arbitrary categories (e.g., random coin toss or badge color; Tajfel et al., 1971). Moreover, it follows that the categorizations used

by researchers to study groups can themselves feed forward, propagating and reinforcing group differences (Gillespie et al., 2012). Similarly, administrative categories can also loop. For example, in the 1970s, the administrative label of 'Hispanic' brought together distinct groups from Latin America and the Iberian Peninsula within a single label, which in turn forged a new political identity for those described by it (Mora, 2014). Yet, it is rare for either administrators or psychologists to reflect on the downstream consequences of categorizations.

Looping theories

While Hacking (1995, 1999) narrowly examined how phenomena interact with classifications, we propose that looping effects are more widespread. Specifically, we argue that many psychological theories are also subject to looping effects but perhaps with less capricious outcomes.

Consider beliefs about willpower. People who believe their willpower is unlimited tend to have more willpower than those who believe it is limited (Francis et al., 2024). Interestingly, people who believe their willpower is a depleting resource are also better able to restore it with glucose (Job et al., 2013). Relatedly, believing in free will also impacts people's behavior, making people behave more ethically (Vohs & Schooler, 2008). Thus, lay theories about willpower and free will can influence these phenomena.

The theory of nudging suffers from destabilizing looping effects. Nudging examines how small changes to choice architectures can prompt desired actions. However, 15% of nudges backfire and have the opposite of the intended effect (Mertens et al., 2022). One explanation is that people who know they are being nudged feel manipulated and resist it. One possible way to address this is to use a participatory nudge, termed nudge+, which entails sharing the details of the nudge with participants. Nudge+ does not create backfire effects and can even improve nudge efficacy (Banerjee et al., 2023).

In addition to being resisted, looping psychological theories can sometimes be repurposed. For example, social identity theory (Brown, 2000), which aims to improve intergroup cohesion, has been used to argue for segregation. Specifically, it has been used to argue that different cultural categories of people don't mix well (Hopkins et al., 1997). In this new racism, there is no denigration or accusations of inferiority. Instead, social psychological research is used to argue that intergroup conflict is an inevitable byproduct of mixing cultural groups. Accordingly, the locus of intergroup tension moves from denigration (theorized by social identity theory) to beliefs about cultural incompatibility (which, in turn, requires novel theorization).

Although looping effects are clearly evident in our categories and theories, there has been little theorization of how they arise and their implications. To date, they have been little more than embarrassing anomalies. But looping effects have become too widespread to ignore. Trying to prevent our categories and theories from becoming representations in society is untenable. Instead, we need to bring looping effects into our conceptualization of what psychological research is. How can we move from being buffeted by unintended looping effects towards harnessing them?

Epistemology of looping effects

Rorty (1981) argued that naive realism has been seduced by the simple but misleading metaphor of theory as a 'mirror' of nature. This metaphor suggests that a good theory is like a faithful mirror that 'captures' perfectly the world before it, reflecting its truth. While this metaphor is often suitable for inanimate objects, when humans see their reflection in a mirror they tend to modify their appearance.

Rorty (1981, p. 48) provocatively argued that theories have "no more of a representational relation to an intrinsic nature of things than does the anteater's snout." The anteater's snout is an adaptation that mediates between the anteater's hunger and nutritious ants in hard-to-reach crevices. Despite being effective, the snout is not a mirror of nature. Similarly, theories about criminality, mental illness, or intergroup relations are less a mirror of nature than ways of acting within nature (e.g., preventing recidivism, treating depression, preventing intergroup conflict). That is to say, *theories of human behavior are meant to transform human behavior*.

The pragmatist approach is to conceptualize theories as 'tools' (Gillespie et al., 2024). Tools do not stand apart from nature but rather operate within nature. Tools are growths within nature, opening up new possibilities and expanding nature itself (Mead, 1932). Substantial psychological theories create new capabilities (e.g., to self-regulate, remember, or influence). Just as the anteater's snout loops into the world, not just consuming ants but also prompting ants to hide in deeper crevices, our psychological theories loop into the world, affecting people. From this pragmatist standpoint, looping effects are not anomalies; they are the expected outcomes of useful theories.

Our approach is to bring theory down from the ethereal plane of timeless knowledge into the practical domain of human activity. Theories not only conceptualize the causes of behavior; *theories can become causes of behavior*. This paradigmatic shift raises new questions. Instead of asking whether knowledge is true or false, we need to ask what its consequences are. Instead of looping effects being an anomaly, we need to ask how we can harness these effects.

Varieties of looping effects

Since Hacking's (1995, 1999) identification of looping effects, it has become apparent that there are a variety of looping effects. Looping effects vary in terms of degree of awareness, domain dependency, and the extent of epistemological destabilization.

Kuorikoski and Pöyhönen (2012) have identified some looping effects that bypass human awareness. Specifically, they argue that categories can be reinforced through social institutions, even without humans being aware of it. For example, knowledge about education, rehabilitation, and incentive structures guides schools, prisons, and corporate environments, which in turn guide these human behaviors. This contrasts with the typical cases of looping described by Hacking (1995, 1999) that operate via human awareness (e.g., a patient becoming aware of diagnostic criteria). Although this looping via institutions underscores the prevalence of looping effects, they are beyond our current scope.

Table 1. Varieties of Looping Effects in Psychology.

	Weakened by looping	Strengthened by looping
Unintentional (accidental)	Deception cues	Clinical therapies (e.g., CBT)
	Priming	Diagnostic criteria
	Nudging	Heuristics
	Persuasion techniques	Social identity categories
	The Pygmalion effect	Stigmatizing labels
Intentional (planned)	Bystander effect training	Informed placebo effect
	Debiasing training	Liberation psychology
	Inoculation/pre-bunking	Metacognitive strategies
	Naive realism	Mnemonic aids
	Stereotype threat mitigation	Nudge+

Another important idea is that looping effects might vary by domain (Vesterinen, 2021). For example, applying a concept from culture X to culture X would produce more congruent looping effects than applying it to culture Z, which might produce more destabilizing looping effects. Such variability reveals how looping effects are deeply cultural and historical, varying across space and time. However, this domain variability does not address our narrower question of why looping can strengthen or weaken theories.

A final variation in looping effects, mentioned by both Kuorikoski and Pöyhönen (2012) and Vesterinen (2021), is that they can be more or less destabilizing. We develop this idea to distinguish between strengthening and weakening effects. Also, while previous literature has focused on the looping effects of categorizations, we focus more broadly on looping theories. Finally, while the previous literature has been more focused on the epistemological problem of unstable categories, our focus is more practical: how can we build theory that incorporates looping effects?

To this end, Table 1 introduces a two-by-two typology to conceptualize two key dimensions of looping effects: weakening or strengthening and unintended or intended. We review each quadrant in turn.

Unintended weakening looping effects

The most prototypical looping effects accidentally weaken theories. For example, priming, which is widely used in psychology to activate subconscious schemas (e.g., priming poverty or old age), often suffers from contrast effects when participants become aware of the priming (Doyen et al., 2012; Lombardi et al., 1987; Molden, 2014). Similarly, if people suspect they are being nudged, the intervention can backfire (Mertens et al., 2022). If people are aware of a persuasion tactic, then they can resist it (Friestad & Wright, 1994). If deceivers are aware of the cues to deception, they can avoid them, thus nullifying the validity of the cues (Levine, 2019). Finally, when teachers and leaders were told about the Pygmalion effect (expecting high performance in others improves their progress), it accidentally attenuated the phenomenon (White & Locke, 2000).

In unintended weakening loops, the instigating theory was made without any consideration of looping effects. Accordingly, the looping results in an accidental attenuation of

the phenomenon being conceptualized. Although this is often viewed as problematic, as we will see, it can be used to weaken undesirable psychological tendencies.

Unintended strengthening looping effects

Unintended strengthening looping effects pertain to theories being reinforced when people become aware of them. For example, labeling someone as deviant may exacerbate their deviant behavior (Becker, 1963). Patients becoming aware of diagnostic categories and criteria may amplify relevant symptoms during a diagnosis (Hacking, 1995, 1999). Patients in counseling may also learn to evaluate their cognitions from a therapist's standpoint, thus amplifying the efficacy of the therapy (Kay et al., 2024). More broadly, it is likely that when clients internalize therapeutic theories and techniques, they become more effective. For example, in cognitive behavior therapy, if clients become aware of the techniques, then there is a boost to the efficacy of the interventions (Beck, 2020).

It is plausible that strengthening loops occur in more domains than is currently recognized, including, for example, conversation analysis (making people better at turn taking, demonstrating, and repairing), well-being research (improving people's happiness), and distributed cognition (enabling people to offload cognition into the physical and social world). In any case, unintended strengthening loops are defined by the instigating theories focusing on a psychological phenomenon, not the potential for looping effects.

Intended weakening looping effects

Looping effects have been used intentionally to weaken psychological processes. For example, when participants were told about the anchoring bias (first impressions guide judgments), the fundamental attribution error (focusing on personality rather than context), and the bias blind spot (noticing biases in others but not oneself), the biases were reduced (Morewedge et al., 2015; but for conflicting evidence, see Pronin et al., 2002). Also, telling people about the naive realism bias (assuming one's own worldview is correct) can reduce the bias (López-Rodríguez et al., 2022). Similarly, graduate students who were told about confirmation bias (favoring confirmatory information) were 29% less likely to choose an inferior hypothesis-confirming course of action (Sellier et al., 2019). Finally, stereotype threat (the tendency to conform to a negative stereotype) can be weakened when people are made aware of the tendency (Johns et al., 2005).

Many biases can be weakened or even extinguished when the theory is looped back to participants. This reveals the potentially transformative impact of theories on human cognition and action. Naming a behavior or cognition, providing a language to talk about it, enables people to see it in themselves and others and thus to decide whether to accept it or resist it. To ignore these looping effects is to blunt our understanding of an important societal contribution.

Intended strengthening looping effects

Looping effects have been used to strengthen desirable patterns of thought and behavior. From liberation psychology (Freire, 1970) to self-help books (Schueller & Parks, 2014),

psychology routinely promises transformation through awareness. Gratitude and mindfulness interventions entail telling people about the psychological theories and practices of gratitude and mindfulness (Henriksen et al., 2020). Similarly, teaching people about metacognitive strategies (reflecting on one's own thinking processes) can improve cognitive performance on a range of tasks (Rivas et al., 2022), including emotion regulation (Denny, 2020). Although human memory has a biological basis, it can be greatly extended with mnemonic aids (Vygotsky & Luria, 1994). For example, learning mnemonic strategies (memory palaces, visceral images, or forming associations) has enabled orators, actors, and stage performers to achieve remarkable feats of memory (Yates, 1974). Heuristics can improve cognitive function beyond memory: in time-pressed situations with imperfect knowledge, fast and frugal heuristics can be effective for decision making (Gigerenzer & Gaissmaier, 2011). Specifically, training people about heuristics and their context-appropriate use can enhance decision making in various contexts (Bingham & Eisenhardt, 2011), including medical practice (Feufel & Flach, 2019).

Across these psychological theories, from metacognition to therapy, looping effects can strengthen psychological theories by making the underlying phenomena more prevalent and effective. The key point is not that people use metacognitive strategies, heuristics, or therapy techniques; it is that their use can be improved through teaching. Learning about these theories strengthens them because, in so far as they are useful, the theories spread and refine the underlying psychological phenomena.

Conceptualizing intended vs. intended loops

One trend revealed by our typology is that theories often migrate from unintended to intended looping effects. These theories began as observations of patterns, and then unintentional looping grew into intentional looping to deliberately alter the phenomenon. This can be done to either weaken (e.g., cognitive biases, stereotype threat) or strengthen the effect (e.g., training in metacognitive strategies, self-regulation). Each of these literatures has discovered the power of looping and incorporated it into research and theorizing.

Consider the bystander effect, namely reduced helping when others are present. It was originally postulated as a law-like rule of human behavior without any consideration of looping (Darley & Latané, 1968). However, the shocking nature of the findings, and the widely discussed case of Kitty Genovese (which was overstated; Manning et al., 2007), meant that it rapidly looped into society. It did not take long before researchers examined the effect of this looping and found that intentionally telling students about the theory weakened the theory (Beaman et al., 1978).

The placebo effect demonstrates how a theory can move from unintended weakening to intended strengthening. Initially, placebo effects were assumed to work only if patients were not aware that they were taking a sugar pill (De Craen et al., 1999). However, once the concept of the placebo effect circulated in society, studies that overtly introduced sugar pills along with an explanation of the power of placebo effects found that sugar pills, even with full awareness, were still effective (von Wernsdorff et al., 2021). Thus, what began as manipulation (patients not being aware of the placebo), with a fear that it would be weakened by looping, was transformed when the theory of the placebo effect

itself was looped (without any manipulation of patients), thus resulting in the effect being strengthened.

Conceptualizing strengthening vs. weakening loops

Underlying these looping effects is the target of the theory becoming aware of the theory. Awareness enables reflexivity and agency (Gillespie, 2018). Awareness gives people choices: to embrace it or resist it; to use the theory to guide action or to ignore it; to create institutions to cultivate it or guard against it; to seek training in the theory or to engage in post-hoc corrections for the behavior. Psychological theories are semiotic mediations (Valsiner, 2001). They give people words for phenomena, thus opening them up to conscious deliberation and guidance. While this explains how awareness instigates looping effects, it does not explain why certain theories are embraced and others resisted.

Psychological reactance theory proposes that people resist theories when they threaten their freedom (Reynolds-Tylus, 2019). If a theory undermines people's autonomy, it may lead to feelings of anger and motivate people to restore their sense of agency by reacting in the opposite way. Such reactance can explain 'backfire' and 'boomerang' effects (Banerjee et al., 2023).

A key issue is who is doing what to whom (i.e., power). Theories that are weakened by looping tend towards controlling people, undermining their autonomy, and making them predictable: the interests of those using the theories are not the same as those targeted by the theories (e.g., persuading, nudging, deceiving). In contrast, theories that are strengthened by looping tend towards empowering people, increasing their capacity to act (e.g., therapy, heuristics, learning). In strengthening loops, there seems to be a closer alignment between the theory users and the people on the receiving end; indeed, they are often the same (e.g., people using the theory on themselves to enhance metacognition or memory).

One lens to conceptualize the strengthening versus weakening effects is Buber's (1937) distinction between I-It and I-Thou relations. I-It relations are prototypically with material things, where resources are instrumentalized, used, and exploited. However, I-It relations can also occur between people, such as when relations are transactional or exploitative. I-Thou relationships are prototypically with other humans, where there is mutual respect, striving to understand and empathize, and a sharing of interests. However, I-Thou relations can also occur with nature (e.g., treating nature or things as sentient or spiritual). Buber argues that one of the defining differences between these modes of relating is the extent to which the other is considered to be known (i.e., closed, predictable, and not worth listening to) or unknown (i.e., open, unpredictable, and worth listening to). Genuine I-Thou relationships must assume that the other can never be fully known, is a source of knowing, and thus is worth listening to.

Psychology approaches people in both modes: sometimes, the aim is to predict and control people (I-It), and sometimes, the aim is to understand, support, and empower people (I-Thou). This gets at a core issue for psychology (Jones & Nisbett, 1972): Does understanding stem from bypassing people's own accounts of their behavior (e.g., a statistical prediction or experimental manipulation), or does it stem from understanding actors' own accounts of their behavior (e.g., engaging with their motivations

Table 2. Conceptualizing Strengthening and Weakening Loops.

	Weakened by looping	Strengthened by looping
Relationship	I-It	I-Thou
Characteristics	People are subjected to psychology	People are given psychology
	Applied to people	Created with people
	Often misaligned goals	Alignment of goals
	Manipulations	Suggestions
	More likely to backfire	Less likely to backfire
	Potentially chasing a moving target	Feeding into an emerging phenomenon
Methods	The research question is withheld	The research question is shared
	Studies 'subjects'	Studies 'participants'
	Undermines subjects' agency	Amplifies participants' agency
	Interpretations about subjects	Interpretations with participants

and reasons)? And what type of understanding should psychology strive to attain? Both types are possible.

Table 2 summarizes the main differences between weakening and strengthening looping effects. Conceptualizing theories as tools, we suggest that strengthening and weakening effects hinge upon who is using the tool, on whom, and for what purpose. When the categories and theories are applied to people, then it is an I-It relationship (i.e., participants are the objects of the tools, not the users of the tools). These relationships, we propose, will often lead to backfire effects that weaken the theories because awareness triggers reactance. However, when the categories and theories are given to participants and used by them, it is an I-Thou relationship (i.e., participants are involved in creating and using the tools). We propose that looping can strengthen the theories in these I-Thou relationships, as the theories are embraced because they are empowering.

In terms of methods, both qualitative and quantitative methods can foster knowledge that strengthens and weakens through looping. The key issue is not the methodology but the asymmetry between the researcher and the people studied. If the researcher withholds the research question and makes interpretations about participants that they would feel uncomfortable sharing with them (i.e., is undermining or irrelevant; Cornish, 2020), then they are likely to produce knowledge that will be resisted. In contrast, if the research addresses a common concern, producing knowledge that is useful and openly shared, then it is more likely to be strengthened through looping.

Chasing a moving target

Another trend revealed by our typology is that theories that experience unintended weakening looping effects can end up chasing a moving target. If a theory only works when there is a knowledge asymmetry and there is looping that erases this asymmetry, then a new theory will likely be needed to reestablish the asymmetry (only to be weakened by subsequent looping).

Consider the case of deception. Initially psychologists identified various non-verbal cues to deception (e.g., looking away, hesitating; Levine, 2019). However, these cues

looped into the hands of deceivers, becoming aids to deception (i.e., avoiding displaying the cues). Thus psychologists had to focus on less controllable cues such as microexpressions, namely brief involuntary facial expressions (Ekman, 1993). However, as the idea of microexpressions has looped into society, people have become increasingly aware of them and more able to inhibit their display (e.g., mindfulness or distraction; Burgoon, 2018). This, in turn, has prompted psychologists to use algorithms on high-resolution videos to identify patterns of increasingly subtle cues that are assumed (for now) to be beyond the awareness of deceivers (Constâncio et al., 2023). Thus, searching for cues to deception is like chasing a moving target. Knowledge to detect lying at time one is used to enable lying at time two in an ongoing loop, thus demonstrating the dynamic role of looping effects in constructing the phenomenon.

Theories of persuasion are also chasing a phenomenon being changed by the theories themselves. The public has increasingly subtle lay theories of who is trying to persuade them, why, and how. Friestad and Wright (1994) describe people as having knowledge not only about persuading agents but also about their persuasion tactics (e.g., advertisements, role modeling, pricing strategies). Of course, the agents of persuasion have adopted increasingly subtle tactics (e.g., influencers, product placements), which have again filtered into common sense. Research has demonstrated how this looping can undermine the tactics. For example, Roozenbeek et al. (2022) showed participants videos about misinformation tactics (emotionally manipulative language, incoherence, false dichotomies, scapegoating, and ad hominem attacks). These videos increased people's ability to identify and ignore these tactics. Continually updating lay knowledge about such tactics means that the agents of persuasion are chasing a moving target. More precisely, each time an effective theory is developed, it circulates and is used by the target audience, whose behaviors become outside the theory.

Psychological knowledge that aims to predict or control people risks becoming dated—as people become aware of it, they react against it. Indeed, as Gergen (1973) observed, the more effective the knowledge, the more vociferous the looping effect. Accordingly, this type of knowledge must keep ahead of common sense while continuously being consumed by common sense. Such a science would not be futile; it would be actively contributing to common sense, transforming the phenomena it theorizes. However, ignoring this looping would make the whole dynamic process an anomaly.

Implications

Imposing a natural science framework on phenomena that are within the domain of *Völkerpsychologie* creates embarrassing anomalies, is theoretically untenable, and weakens theories because a persistent source of variance is being overlooked. In many domains, our theories are causes of behavior, which in turn makes the theories epistemologically unstable. How can we move beyond being victims of these looping effects towards conceptualizing them and incorporating them into our theories?

There is a brewing crisis of theory in psychology. Muthukrishna and Henrich (2019) argue that psychology is missing an overarching theoretical framework to situate findings. Eronen and Bringmann (2021) argue that psychological theory has been held back by a lack of constraints, measurement validity, and obstacles to identifying causes. We

agree with these assessments and contribute an additional obstacle for psychological theory: looping effects. Looping effects are so pervasive that they need to be brought into the frame for any comprehensive psychological theory. The attempt to create theories without considering looping effects has led to theories that risk being relegated to history by being folded into common sense. No amount of methodological rigor will overcome the challenge of looping effects, *because it is a theoretical challenge*.

Psychological theory aims to identify long-lasting and useful causal relationships. However, it will not achieve this goal by imposing a natural science conception of theory onto the domains of *Völkerpsychologie*; this risks backfire effects and weakened theory. Rather than viewing looping effects as an anomalous taboo, we propose to bring looping effects into theory construction. Looping effects are a powerful cause of human behavior that has been hiding in plain sight.

Bringing looping effects into psychological theory entails a paradigmatic shift towards theorizing the role of psychology in individual, cultural, and societal development. Instead of theorizing what humans are or how they develop, the question shifts to how psychology can augment humans (i.e., the role of psychology in human development). To this end, our theories must be reconceptualized within a feedback loop (Kuorikoski & Pöyhönen, 2012). The theory and the phenomenon develop together, with each shaping the other. Observed behavior shapes the theories, the theories shape the behavior, which in turn shapes the theories, and so on. Not all phenomena are equally malleable: at one extreme, the theory has minimal capacity to shape the phenomenon (e.g., natural science phenomena, such as genetics, mirror neurons, spike potentials), while at the other extreme, the theory has maximal capacity to shape the phenomenon (e.g., social science phenomena, such as norms, inequalities, and even economic theories). This theory-phenomenon dynamic system is constrained by both the underlying phenomena and the power of our own theories. Nevertheless, the theory and the phenomenon, together and over time, are charting a route through the space of human potential.

There are several concrete steps that can be taken to bring looping effects into our theories. When testing hypotheses, we should specify not only how we expect people to behave under this or that condition but also how we expect them to behave when they learn about the theory. Would looping lead to strengthening or weakening? In experiments, a condition could be added where participants are told in advance about their expected behavior. Or, it can be added as a control. For example, in the failed replication of the hypothesis that inducing a smile (by holding a pen in one's teeth) makes cartoons more funny, the researchers controlled for awareness of the hypothesis (Coles et al., 2022). They found that many participants were indeed aware of the hypothesis, and, moreover, the behavior of aware participants was more in line with the hypothesis.

In survey research, a question could be included to identify participants who were aware of the guiding theory. Equally, survey questions could identify participants who assumed a different or even inaccurate theory. However, perhaps most important for survey research, which emphasizes representativeness, is to understand how looping effects interact with growing non-response rates (Groves, 2011). In clinical domains, the criteria for diagnosis could include expectations for how symptoms might interact with

the client's awareness of the symptoms (e.g., whether certain symptoms tend to be accentuated or suppressed). But, of course, if it was widely known that awareness of the diagnostic criteria was part of the diagnostic criteria, it might lead to second-level looping (e.g., hiding awareness). In qualitative research, there could be an explicit discussion of what participants thought about the guiding research questions, theory, and motives. More generally, there should be continual reflection, not just on the downstream consequences of our theories but also on how the theories themselves interact with their phenomena (Brinkmann, 2005). Over time, these steps would lead to a more nuanced meta-theory of how our theories guide human behavior.

Looping effects are often small, but so are many effects in psychology (Eronen & Bringmann, 2021; Meehl, 1990). In the context of the replication and theory crises, we cannot afford to ignore the variance produced by looping effects because it is theoretically inconvenient. We need to recognize that large parts of psychology are in the domain of *Völkerpsychologie*, and thus subject to looping effects. Until then, our theories will continue to be challenged by inconsistent findings, unexplained variance, and propositions that fail to transcend their historical context.

Limitations

The proposed analysis has limitations. The typology of looping effects does not categorize all theories easily. We have already drawn attention to theories that began with unintended looping effects and have had increasingly intentional effects (e.g., the bystander effect, the Pygmalion effect). Equally, there are some theories that can provoke both weakening and strengthening, making them difficult to place (e.g., labeling deviance; Becker, 1963). Whether a theory is strengthened or weakened will depend, in part, on participants' interpretation of the aims of the theory (e.g., to empower them or to control them), and this will have a historical aspect (e.g., changing norms).

The main limitation is that not all theories have looping effects, and not all looping effects are particularly consequential (Tsou, 2007). Despite widespread publicity, some social phenomena have not weakened as we might expect (e.g., obedience; Doliński et al., 2017). Also, no amount of looping (in the medium term) will substantially alter how our eyes process images, our grammatical use of a language, or the foundational role of metaphors in cognition. Additionally, there are parts of psychology that belong in the natural sciences (e.g., genetics, mirror neurons, evolution by natural selection). However, future choices may arise as our knowledge in these domains advances (e.g., gene manipulation). However, even choices about genetic manipulation won't change the underlying mechanisms that lead from the double helix to a living organism. Our theories are far from enabling such actions.

One approach is to distinguish fundamental psychological theory (identifying underlying causes) from the 'soft' side of psychology that focuses on the downstream manifestations of these theories (Meehl, 1990)—essentially reasserting the line between where natural science ends and *Völkerpsychologie* begins. However, this line is difficult to draw because as knowledge grows, so does our capacity to intervene in our own psychology and behavior; it is a moving boundary. Yet, there is a distinction to be made: Looping effects have limits, and it is necessary to identify where they currently are.

Conclusion: The aims of psychology

Looping effects in psychology are pervasive. Our categories and theories become part of society, mediating human behavior and cognition. The effects can vary from backfiring to zealous uptake. Some theories are weakened or undermined, while others are strengthened. Nevertheless, there is an increasing trend to take control of these looping effects. Originally postulated as rule-like, psychological theories have increasingly been taught so as to selectively strengthen and weaken psychological tendencies.

Acknowledging that psychological categories and theories are part of society (and not outside it) means that looping effects need to be built into psychological theories. It means moving away from the metaphor of 'the mirror of nature' and instead conceptualizing psychological knowledge as tools for action. The key question becomes: Who is enabled by our tools to do what, and to whom?

If the aim of psychology is to predict, control, or otherwise manipulate the public, then looping effects will be a continual problem. One could try to prevent looping by selectively ceasing to circulate effective theories. Despite being unethical and unfeasible in an open society, this would also run counter to psychology's ambition to contribute to the betterment of society and to 'give psychology away' (Miller, 1969, p. 1071). Alternatively, the discipline could skew towards natural science phenomena, which are less exposed to looping effects. However, this is untenable because it would entail withdrawing from the social science end of the discipline, which is where much of the societal impact of psychology occurs (Zimbardo, 2004). Finally, the discipline could continue to impose a natural science paradigm on the social side of the discipline. However, this will continue to produce anomalous looping effects. In so far as psychological theory is in the service of third parties, enabling them to predict and control their targets, we foresee an endless treadmill of theory creation and subsequent backfire effects.

If psychology aims to make empowering theories (or tools) for people to address their own concerns and become more agentic, then we foresee strengthening effects. Aiming to create knowledge that is useful for laypeople will push towards increasing goal alignment between researchers and the public, which in turn will make psychological research more relevant and more significant for society. It would shift the focus from fundamental research to applied research (Andersen, 2025), with extra benefits. By acknowledging the historicity of psychological knowledge and bringing participants into the research process, by aligning our interests with theirs, we are more likely to create strengthening looping effects that will, paradoxically, make psychological knowledge more robust in the face of history.

The gentleman with arms akimbo, described by Dostoevsky (1864) in our opening quote, is resisting being predicted by mathematical theories and algorithms. He captures the reactionary attitude that most humans have towards being predicted, controlled, or manipulated. Few human identities can tolerate being viewed as puppets animated by a formula. If the predicted behavior is under voluntary control, then it is likely that awareness will lead to reactance. However, such reactionary responses, instead of being conceptualized as an embarrassing bug, can be reconceptualized as revealing the historical, cultural, and continually self-developing essence of humans. We should celebrate the cases when psychology has contributed to weakening undesirable behaviors and

strengthening desirable ones. It shows how theories about humans inevitably fold into what humans become. This looping, we argue, does not have to be an anomaly; it can be brought within psychological theory, leading to more robust theory, less temperamental findings, and increased societal impact.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Alex Gillespie would like to acknowledge the NCCR-On the Move, funded by the Swiss National Science Foundation under grant number 51NF40-205605.

ORCID iD


Alex Gillespie https://orcid.org/0000-0002-0162-1269

References

Andersen, J. P. (2025). Applied research is the path to legitimacy in psychological science. *Nature Reviews Psychology*, *4*(1), 1–2. https://doi.org/10.1038/s44159-024-00388-9

Arnett, J. J. (2016). The neglected 95%: Why American psychology needs to become less American. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 115–132). American Psychological Association.

Banerjee, S., Galizzi, M. M., John, P., & Mourato, S. (2023). Immediate backfire? Nudging sustainable food choices and psychological reactance. *Food Quality and Preference*, *109*, Article 104923. https://doi.org/10.1016/j.foodqual.2023.104923

Baucal, A., Gillespie, A., Krstić, K., & Zittoun, T. (2020). Reproducibility in psychology: Theoretical distinction of different types of replications. *Integrative Psychological and Behavioral Science*, *54*(1), 152–157. https://doi.org/10.1007/s12124-019-09499-y

Baumeister, R. F., Tice, D. M., & Bushman, B. J. (2023). A review of multisite replication projects in social psychology: Is it viable to sustain any confidence in social psychology's knowledge base? *Perspectives on Psychological Science*, *18*(4), 912–935. https://doi.org/10.1177/17456916221121815

Beaman, A. L., Barnes, P. J., Klentz, B., & McQuirk, B. (1978). Increasing helping rates through information dissemination: Teaching pays. *Personality and Social Psychology Bulletin*, *4*(3), 406–411. https://doi.org/10.1177/014616727800400309

Beck, J. S. (2020). *Cognitive behavior therapy: Basics and beyond*. Guilford Publications.

Becker, H. S. (1963). *Outsiders: Studies in the sociology of deviance*. The Free Press.

Bingham, C. B., & Eisenhardt, K. M. (2011). Rational heuristics: The 'simple rules' that strategists learn from process experience. *Strategic Management Journal*, *32*(13), 1437–1464. https://doi.org/10.1002/smj.965

Blank, T. O. (1988). Reflections on Gergen's "social psychology as history" in perspective. *Personality and Social Psychology Bulletin*, *14*(4), 651–663. https://doi.org/10.1177/0146167288144001

Brinkmann, S. (2005). Human kinds and looping effects in psychology: Foucauldian and hermeneutic perspectives. *Theory & Psychology*, *15*(6), 769–791. https://doi.org/10.1177/0959354305059332

Brinkmann, S. (2016). *Diagnostic cultures: A cultural approach to the pathologization of modern life*. Routledge.

Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. *European Journal of Social Psychology*, *30*(6), 745–778. https://doi.org/10.1002/1099-0992(200011/12)30:6%3C745::AID-EJSP24%3E3.0.CO;2-O

Buber, M. (1937). *I and Thou* (R. G. Smith, Trans.). T & T Clark.

Burgoon, J. K. (2018). Microexpressions are not the best way to catch a liar. *Frontiers in Psychology*, *9*, 1672. https://doi.org/10.3389/fpsyg.2018.01672

Chomsky, N. (1995). *The minimalist program*. Cambridge University Press.

Coles, N. A., March, D. S., Marmolejo-Ramos, F., Larsen, J. T., Arinze, N. C., Ndukaihe, I. L. G., Willis, M. L., Foroni, F., Reggev, N., Mokady, A., Forscher, P. S., Hunter, J. F., Kaminski, G., Yüvrük, E., Kapucu, A., Nagy, T., Hajdu, N., Tejada, J., Freitag, R. M. K., . . . Liuzza, M. T. (2022). A multi-lab test of the facial feedback hypothesis by the Many Smiles Collaboration. *Nature Human Behaviour*, *6*(12), 1731–1742. https://doi.org/10.1038/s41562-022-01458-9

Constâncio, A. S., Tsunoda, D. F., de Fátima Nunes Silva, H., da Silveira, J. M., & Carvalho, D. R. (2023). Deception detection with machine learning: A systematic review and statistical analysis. *PLoS One*, *18*(2), e0281323. https://doi.org/10.1371/journal.pone.0281323

Cornish, F. (2020). Communicative generalisation: Dialogical means of advancing knowledge through a case study of an 'unprecedented' disaster. *Culture & Psychology*, *26*(1), 78–95. https://doi.org/10.1177/1354067X19894930

Corrigan, P. W. (2016). Lessons learned from unintended consequences about erasing the stigma of mental illness. *World Psychiatry*, *15*(1), 67–73. https://doi.org/10.1002/wps.20295

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge University Press.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4), 377–383. https://doi.org/10.1037/h0025589

De Craen, A. J. M., Kaptchuk, T. J., Tijssen, J. G. P., & Kleijnen, J. (1999). Placebos and placebo effects in medicine: Historical overview. *Journal of the Royal Society of Medicine*, *92*(10), 511–515. https://doi.org/10.1177/014107689909201005

Denny, B. T. (2020). Getting better over time: A framework for examining the impact of emotion regulation training. *Emotion*, *20*(1), 110–114. https://doi.org/10.1037/emo0000641

Descartes, R. (1641). Meditations on first philosophy. In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical writings of Descartes (Volume II)* (pp. 2–63). Cambridge University Press.

Doliński, D., Grzyb, T., Folwarczny, M., Grzybała, P., Krzyszycha, K., Martynowska, K., & Trojanowski, J. (2017). Would you deliver an electric shock in 2015? Obedience in the experimental paradigm developed by Stanley Milgram in the 50 years following the original studies. *Social Psychological and Personality Science*, *8*(8), 927–933. https://doi.org/10.1177/1948550617693060

Dostoevsky, F. (1864). *Notes from the underground and the double*. Penguin.

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS One*, *7*(1), Article e29081. https://doi.org/10.1371/journal.pone.0029081

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392. https://doi.org/10.1037/0003-066X.48.4.384

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, *16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Farr, R. M. (1983). Wilhelm Wundt (1832-1920) and the origins of psychology as an experimental and social science. *British Journal of Social Psychology*, *22*(4), 289–301. https://doi.org/10.1111/j.2044-8309.1983.tb00596.x

Feufel, M. A., & Flach, J. M. (2019). Medical education should teach heuristics rather than train them away. *Medical Education*, *53*(4), 334–344. https://doi.org/10.1111/medu.13789

Fodor, J. (1983). *The modularity of mind*. MIT Press.

Francis, Z., Weidmann, R., Bühler, J. L., Burriss, R. P., Wünsche, J., Grob, A., & Job, V. (2024). My willpower belief and yours: Investigating dyadic associations between willpower beliefs, social support, and relationship satisfaction in couples. *European Journal of Personality*, *38*(5), 778–792. https://doi.org/10.1177/08902070231220416

Freire, P. (1970). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). Continuum.

Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, *21*(1), 1–31. https://doi.org/10.1086/209380

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*, 309–320. https://doi.org/10.1037/h0034436

Gergen, K. J. (2015). From mirroring to world-making: Research as future forming. *Journal for the Theory of Social Behaviour*, *45*(3), 287–310. https://doi.org/10.1111/jtsb.12075

Gerlach, P. (2017). The games economists play: Why economics students behave more selfishly than other students. *PLoS One*, *12*(9), Article e0183814. https://doi.org/10.1371/journal.pone.0183814

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gillespie, A. (2018). Distinguishing two processes of self-reflection. In J. Valsiner & A. Rosa (Eds.), *The Cambridge handbook of sociocultural psychology* (pp. 245–259). Cambridge University Press.

Gillespie, A., Glăveanu, V., & de Saint Laurent, C. (2024). *Pragmatism and methodology: Doing research that matters with mixed methods*. Cambridge University Press.

Gillespie, A., Howarth, C., & Cornish, F. (2012). Four problems for researchers using social categories. *Culture and Psychology*, *18*(3), 391–402. https://doi.org/10.1177/1354067X12446236

Groves, R. M. (2011). Three eras of survey research. *Public Opinion Quarterly*, *75*(5), 861–871. https://doi.org/10.1093/poq/nfr057

Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary approach* (pp. 351–394). Oxford University Press.

Hacking, I. (1999). *The social construction of what?* Harvard University Press.

Heasman, B., & Gillespie, A. (2019). Participants over-estimate how helpful they are in a two-player game scenario toward an artificial confederate that discloses a diagnosis of autism. *Frontiers in Psychology*, *10*, 1349. https://doi.org/10.3389/fpsyg.2019.01349

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henriksen, D., Richardson, C., & Shack, K. (2020). Mindfulness and creativity: Implications for thinking and learning. *Thinking Skills and Creativity*, *37*, Article 100689. https://doi.org/10.1016/j.tsc.2020.100689

Herder, J. G. (1784). *Ideen zur philosophie der geschichte der menschheit* [Ideas for the philosophy of history of humanity]. Hartknoch.

Hopkins, N., Reicher, S., & Levine, M. (1997). On the parallels between social cognition and the "new racism." *British Journal of Social Psychology*, *36*(3), 305–329. https://doi.org/10.1111/j.2044-8309.1997.tb01134.x

Job, V., Walton, G. M., Bernecker, K., & Dweck, C. S. (2013). Beliefs about willpower determine the impact of glucose on self-control. *Proceedings of the National Academy of Sciences*, *110*(37), 14837–14842. https://doi.org/10.1073/pnas.1313475110

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, *16*(3), 175–179. https://doi.org/10.1111/j.0956-7976.2005.00799.x

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, &

B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). General Learning Press.

Kant, I. (1871). *Critique of pure reason* (J. C. Meredith, Trans.). JM Dent London.

Kay, E., Gillespie, A., & Cooper, M. (2024). From conflict and suppression to reflection: Longitudinal analysis of multivoicedness in clients experiencing depression. *Journal of Constructivist Psychology*, *37*(3), 239–258. https://doi.org/10.1080/10720537.2023.2175752

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Kuorikoski, J., & Pöyhönen, S. (2012). Looping kinds and social mechanisms. *Sociological Theory*, *30*(3), 187–205. https://doi.org/10.1177/0735275112457911

Laimann, J. (2020). Capricious kinds. *The British Journal for the Philosophy of Science*, *71*(3), 1043–1068. https://doi.org/10.1093/bjps/axy024

Levine, T. R. (2019). *Duped: Truth-default theory and the social science of lying and deception*. University of Alabama Press.

Lindholm, S. K., & Wickström, A. (2020). 'Looping effects' related to young people's mental health: How young people transform the meaning of psychiatric concepts. *Global Studies of Childhood*, *10*(1), 26–38. https://doi.org/10.1177/2043610619890058

Lombardi, W. J., Higgins, E. T., & Bargh, J. A. (1987). The role of consciousness in priming effects on categorization: Assimilation versus contrast as a function of awareness of the priming task. *Personality and Social Psychology Bulletin*, *13*(3), 411–429. https://doi.org/10.1177/0146167287133009

López-Rodríguez, L., Halperin, E., Vázquez, A., Cuadrado, I., Navas, M., & Gómez, Á. (2022). Awareness of the psychological bias of naïve realism can increase acceptance of cultural differences. *Personality and Social Psychology Bulletin*, *48*(6), 888–900. https://doi.org/10.1177/01461672211027034

Manning, R., Levine, M., & Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of the 38 witnesses. *American Psychologist*, *62*(6), 555–562. https://doi.org/10.1037/0003-066X.62.6.555

Mead, G. H. (1932). *The philosophy of the present*. Open Court.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*(1), 195–244. https://doi.org/10.2466/pr0.1990.66.1.195

Mertens, S., Herberz, M., Hahnel, U. J. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences*, *119*(1), Article e2107346118. https://doi.org/10.1073/pnas.2107346118

Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist*, *24*(12), 1063–1075. https://doi.org/10.1037/h0028988

Molden, D. C. (2014). Understanding priming effects in social psychology: What is "social priming" and how does it occur? *Social Cognition*, *32*(Supplement), 1–11. https://doi.org/10.1521/soco.2014.32.supp.1

Mora, G. C. (2014). *Making Hispanics: How activists, bureaucrats, and media constructed a new American*. University of Chicago Press.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140. https://doi.org/10.1177/2372732215600886

Moscovici, S. (2008). *Psychoanalysis: Its image and its public*. Polity.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Power, S. A., Zittoun, T., Akkerman, S., Wagoner, B., Cabra, M., Cornish, F., Hawlina, H., Heasman, B., Mahendran, K., Psaltis, C., Rajala, A., Veale, A., & Gillespie, A. (2023). Social psychology of and for world-making. *Personality and Social Psychology Review*, *27*(4), 378–392. https://doi.org/10.1177/10888683221145756

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Reynolds-Tylus, T. (2019). Psychological reactance and persuasive health communication: A review of the literature. *Frontiers in Communication*, *4*, Article 56. https://doi.org/10.3389/fcomm.2019.00056

Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology*, *13*, Article 913219. https://doi.org/10.3389/fpsyg.2022.913219

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), Article eabo6254. https://doi.org/10.1126/sciadv.abo6254

Rorty, R. (1981). *Philosophy and the mirror of nature*. Princeton University Press.

Rose, N. (2018). *Our psychiatric future*. Polity.

Sadeghiyeh, H., Wang, S., Kyllo, H. M., Alberhasky, M. R., Savita, S., Kellohen, K. L., & Wilson, R. C. (2021). On the psychology of the psychology subject pool: An exploratory test of the good student effect. *Journal of Individual Differences*, *42*(1), 30–40. https://doi.org/10.1027/1614-0001/a000327

Schueller, S. M., & Parks, A. C. (2014). The science of self-help. *European Psychologist*, *19*(2), 145–155. https://doi.org/10.1027/1016-9040/a000181

Sellier, A.-L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*, *30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tsou, J. Y. (2007). Hacking on the looping effects of psychiatric classifications: What is an interactive and indifferent kind? *International Studies in the Philosophy of Science*, *21*(3), 329–344. https://doi.org/10.1080/02698590701589601

Underhill, R., & Foulkes, L. (2024). Self-diagnosis of mental disorders: A qualitative study of attitudes on Reddit. *Qualitative Health Research*, *35*(7), 779–792. https://doi.org/10.1177/10497323241288785

Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development*, *44*(2–3), 84–97. https://doi.org/10.1159/00005704

Vesterinen, T. (2021). Identifying the explanatory domain of the looping effect: Congruent and incongruent feedback mechanisms of interactive kinds. *Journal of Social Ontology*, *6*(2), 159–185. https://doi.org/10.1515/jso-2020-0015

Vico, G. (1725). *Principi di una Scienza Nuova intorno alla natura delle nazioni* [Principles of a new science concerning the nature of nations]. Domenico Antonio Muzi.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, *19*(1), 49–54. https://doi.org/10.1111/j.1467-9280.2008.02045.x

von Wernsdorff, M., Loef, M., Tuschen-Caffier, B., & Schmidt, S. (2021). Effects of open-label placebos in clinical trials: A systematic review and meta-analysis. *Scientific Reports*, *11*(1), Article 3855. https://doi.org/10.1038/s41598-021-83148-6

Vygotsky, L. S., & Luria, A. (1994). Tool and symbol in child development. In R. Van de Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 99–174). Blackwell.

White, S. S., & Locke, E. A. (2000). Problems with the Pygmalion effect and some proposed solutions. *The Leadership Quarterly*, *11*(3), 389–415. https://doi.org/10.1016/S1048-9843(00)00046-1

Wundt, W. (1912). *Elements of folk psychology: Outlines of a psychological history of the development of mankind*. George Allen and Unwin.

Yates, F. A. (1974). *The art of memory*. University of Chicago Press.

Zimbardo, P. G. (2004). Does psychology make a significant difference in our lives? *American Psychologist*, *59*(5), 339–351. https://doi.org/10.1037/0003-066X.59.5.339

Author biographies

Alex Gillespie is a Professor of Psychological and Behavioural Science at the London School of Economics, a Visiting Professor at the Oslo New University College, and an Editor of *Journal for the Theory of Social Behaviour*. His research examines communication problems, especially speaking-up, defensiveness, misunderstandings, distrust, and problems of listening to and learning from challenging feedback. He recently co-authored 'Pragmatism & Methodology' published by Cambridge University Press that is freely available for download.

Brady Wagoner is Professor of Psychology at the University of Copenhagen, Aalborg University and Oslo New University College. He received his PhD from the University of Cambridge on a Gates Cambridge Scholarship. His research aims to develop a dynamic cultural psychology, which he has applied to such topics as memory, social change, and the public understanding of science. His books include *The Constructive Mind: Bartlett's Psychology in Reconstruction* (Cambridge University Press, 2017) and *Handbook of Culture and Memory* (Oxford University Press, 2018). In 2021, he received the prestigious Humboldt Award from the Alexander von Humboldt Foundation.