Perspectives on Psychological Science, in press

Social psychology's empty-self metaphor and the replication crisis

Running Head: **The Empty-Self Metaphor**

Jack W. Klein[1]*
William B. Swann, Jr[2]

[1]Hong Kong Institute of Asia-Pacific Studies, Chinese University of Hong Kong
[2]Department of Psychology, University of Texas at Austin and University of Oregon

*Corresponding author information: Jack W. Klein, Room 507, 5/F, Esther Lee Building, Hong Kong Institute of Asia-Pacific Studies, The Chinese University of Hong Kong, Hong Kong.

Email: jack.klein@cuhk.edu.hk

Abstract

Since the early 20th century, an emphasis on the causal power of situations in social psychology has fostered the view that the self is an empty vessel filled by the contents of the situation. We label this the empty-self metaphor, with incarnations including situationism and elements of theories of self-presentation, self-perception, social identity, the dramaturgical movement, and others. The persistence of this meta-theoretical assumption has led to an underappreciation of an enduring, unique self and to the development of contemporary paradigms (e.g., social priming and embodied cognition) that have hinged on the implicit premise that the self is empty or passive. The self is not empty, of course, and new preliminary evidence we have collected indicates that research predicated on the empty-self is far less likely to replicate. While emphasizing the power of the situation has yielded important theoretical and practical insights, we propose that the field would be strengthened by better accounting for the chronic, dispositional motivations that emanate from an enduring self. We offer suggestions – both theoretical and methodological – that can help social psychologists achieve this goal.

Keywords: The Self, Situationism, Replication Crisis, Theory Crisis.

Social psychology's empty-self metaphor and the replication crisis

The recent "replication crisis" has rocked the field of social psychology. A host of high-profile findings have failed to replicate, particularly those related to social priming and its social embodiment subfield. To list only a few examples: priming old age did not reduce walking speed (Doyen et al., 2012), plotting points on paper further apart did not make people feel less close to their own family (Pashler et al., 2012), and holding a hot cup of coffee did not make people judge others' personalities as warmer (Chabris et al., 2019). The dominant reaction to these failures to replicate has been to identify and critique methodological and statistical practices — low N studies, p hacking, HARKing, deliberate falsification — that have putatively undermined progress in the field (Schiavone & Vazire, 2022; Shrout & Rodgers, 2018; Wicherts et al., 2016). While acknowledging these methodological/statistical critiques, others have sounded the alarm on the dangers of an independent "theory crisis". This critique suggests that much of social psychology lacks plausible or sufficiently developed theorizing (Eronen & Bringmann, 2021). In this essay, we bring these two critiques together by postulating that replication failures may partly stem from an overreliance on a theoretical premise we have labeled the "empty-self metaphor". This premise underlies many of social psychology's most sensational, counterintuitive, and celebrated findings, which have proved alluring to researchers and incentivized the use of questionable research methods. We suggest that the empty-self metaphor may have hindered the capacity of social psychologists to translate the enormous potential of the field to generate solutions to some of the most vexing challenges facing contemporary society.

We assume that the self – which we define as the unique character of people that includes their enduring self-views, traits, knowledge, and values – is an important determinant of human

behavior. This assumption clashes sharply with the empty-self metaphor of personhood that has guided a large contingent of researchers for most of the last century and, even if rarely explicitly acknowledged, remains influential to this day. Put simply, the empty-self metaphor ignores or downplays the role of the self and asserts or implies that people are empty vessels that are passively or reactively filled by the contents of the situation. Within this framework, subtle environmental cues largely dictate people's cognitions and behavior, despite lying outside our conscious awareness, unknown to all but the perceptive psychologist. We do not seek to downplay the well-documented impact of relevant situational influences on behavior, which has inspired insightful formulations we generally hold in high regard (e.g., the social identity approach). Rather, we contend that a complete understanding of responses to situations requires consideration of an enduring self, and that ignoring the self makes the power of the situation harder to reliably demonstrate.

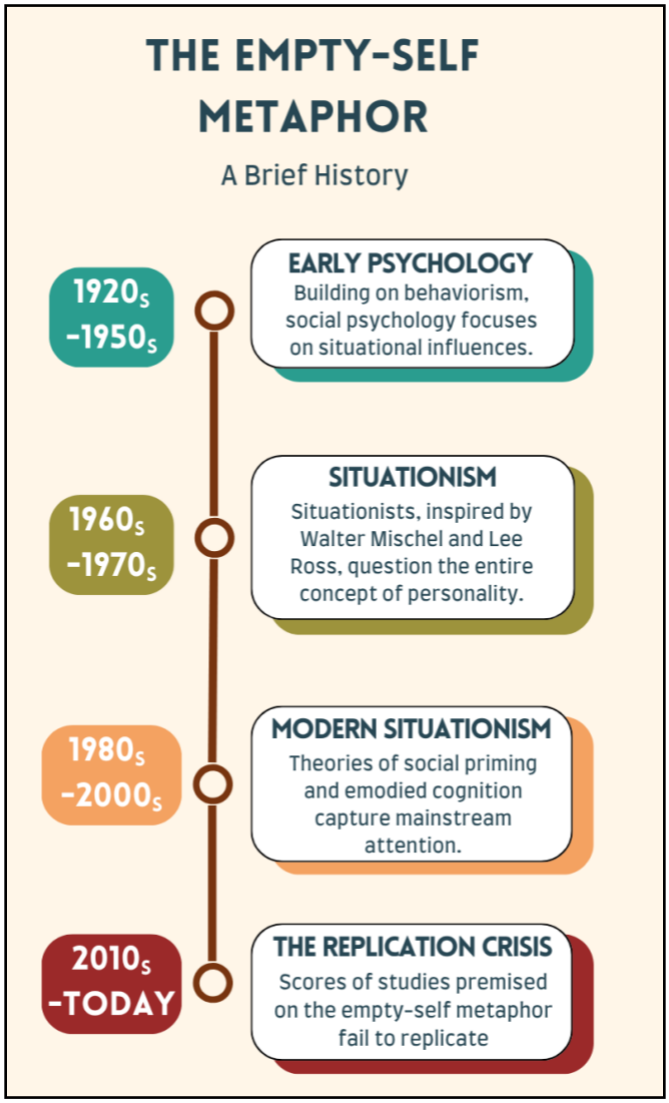
A brief history of the empty-self metaphor

The empty-self metaphor has a long history in psychology (see Figure 1). It arguably had its origins in the behaviorist movement of the early 1900s, which contended that situations are uniquely powerful determinants of behavior and can be experimentally examined more readily than the "black box" of the mind. For these reasons, theorists contended that behavioral scientists should focus their attention on the influence of situations (Watson, 1913). Soon thereafter, behavioristic assumptions made their way into social psychology textbooks (Allport, 1924). Although Floyd Allport, the "father of experimental social psychology" (Katz, 1979), rejected methodological behaviorism (which eschewed all discussion of cognitive and emotional mechanisms), he regarded consciousness as epiphenomenal and thus less worthy of study than situational influences. A few decades later Floyd's brother Gordon enshrined situational

influences into the very definition of social psychology: "how the thought, feeling, and behavior of individuals *are influenced by* the actual, imagined, or implied presence of other human beings" (Allport, 1954 p. 1, italics added).

Figure 1

A timeline of the empty-self metaphor



Emphasis on the power of the situation persisted even when researchers became interested in phenomena that were nominally linked to the self. Festinger's (1957) investigations of attitude change (cognitive dissonance) generally avoided scrutiny of people's firmly held beliefs and attitudes and instead focused on the impact of situational inputs on what were

assumed to be weakly held attitudes. Described as "the psychology of what people do to recover from experimentally engineered major embarrassments" (pg. 43; Abelson, 1983), in its early incantations cognitive dissonance was primarily a theory of impression management, with little regard for an enduring self. It is true that later revisions of dissonance theory acknowledged the importance of self-consistency (Aronson, 1968, 2019). However, most empirical realizations still assumed that the self was effectively empty, with only a few underdeveloped efforts to integrate an enduring self [1]. Bem took this argument further by questioning whether individuals even have access to their internal states and attitudes (1972). To the contrary, his self-perception theory proposed that people must infer their own attitudes from their behavior, just as objective observers must infer the attitudes of others from their behavior[2]. We can see the historical trajectory of the empty-self metaphor when Bem explicitly traces the origins of self-perception theory to the work of radical behaviorists such as B.F. Skinner, who was steadfast in his reluctance to attribute behavior to internal motivations (1972).

Walter Mischel fueled this early situationism with a blistering attack on the notion that stable personality traits guide behavior. His core argument was that because scores on personality scales were poor predictors of behavior across situations and over time, "the entire concept of personality traits as broad dispositions is thus untenable" (Mischel, 1968; p. 146). This implied to many that, by default, behavior must be controlled by the situation. From this vantage point, researchers should suspend efforts to measure an enduring self and shift their attention to

[1] For example, although the occasional study found that a cognitive dissonance response can be triggered by giving participants unexpected positive feedback following a string of negative feedback (e.g., Aronson & Carlsmith, 1962), this paradigm generally failed to replicate. This is probably because it ignores the influence of an enduring self that pre-dates the experimental manipulation, which presumably led most participants to expect positive feedback (Swann, 1990).

[2]Although Bem specified that self-perception processes occurred only when "internal cues were weak, ambiguous, or uninterpretable," researchers generally assumed this to be the case. This aspect of the theory was hence seldom recognized or tested, with few exceptions (e.g., Chaiken & Baldwin, 1981).

situational influences. A decade later, Lee Ross (Ross, 1977) described the very tendency to contemplate the existence of personal or dispositional factors as a "fundamental attribution error". Echoing earlier psychologists (especially Ichheiser, 1943), he warned that the fundamental attribution error was what enabled personality psychology and "erroneous trait inferences and trait theories" to survive (Ross, 1977; p. 187). Although Mischel later distanced himself from the strongest claims of the situationists and advanced an interactionist account (Mischel, 2009), he and Ross continued to cast doubt on a stable self as late as 2016 on an NPR podcast episode titled "The Personality Myth" (Spiegel, 2016).

Although our focus here is on the role of the empty-self metaphor in psychology, and particularly social psychology, we should note it has also penetrated other disciplines. For instance, in sociology, a similar emphasis on situations at the expense of people's enduring sense of self has permeated theory and research on self-presentation. In the tradition of the dramaturgical movement (e.g., Goffman, 1959), researchers assumed that people are like actors in a play who perform for different audiences. Here the self is a consequence rather than a cause of the performance, representing a "product of the scene" (Goffman, 1959, p. 252). People are obligated to remain "in character" until they move to the next scene, at which point they discard the prior self in favor of one that fits the new context. The dramaturgical movement also made room for a "backstage" in which one could freely express their opinions, rehearse, and recover away from the "audience" (Goffman, 1959). Nevertheless, it was the carefully orchestrated "frontstage" in which the "performance" actually occurred, leaving any enduring beliefs existing in the "backstage" as effectively epi-phenomenal, at least in terms of predicting behavior. This vision of the self not only shaped accounts of self-presentation (Baumeister & Hutton, 1987; Jones, 1964; Schlenker, 1980; Tedeschi, 1981), but also encouraged researchers to focus

narrowly on a single goal: understanding how people gain the approval of "the audience" (i.e., other people). Within this framework, people were presumably in the business of constructing whichever identities they believed would help them win the favor of their interaction partners, with the only caveat being that they should avoid the appearance of inconsistency or dishonesty (Schlenker, 1980; Schlenker & Leary, 1985). Allegiance to an enduring self was nowhere to be found in such accounts.

The emptiness of the empty-self metaphor

On the face of it, the empty-self metaphor is clearly alluring. First, the notion that the self is empty introduces the possibility that positive behavior can be elicited via relatively effortless interventions. Consider, for example, the widespread adoption of "nudges" – in which subtly altering the framing of a decision can greatly increase desired behavior (Thaler & Sunstein, 2021) – by both government and business. Second, the idea of imperceptible forces influencing human behavior has reliably captured the public's imagination. There is a reason that the best-selling book *Thinking, Fast and Slow* devoted a chapter to social priming (Kahneman, 2011), that popular mentalists like Derren Brown claim to use cues and primes to control unsuspecting test subjects, and that NBC produced *The Irrational* in which a fictional protagonist (based on social psychologist Dan Ariely) solves crime using his knowledge of unconscious behavior. It is unsurprising that social psychologists would rush to work on topics that are celebrated in popular media, and feel pressured to use questionable research methods to obtain these desirable results.

There are also reasons why social psychologists specifically have embraced the empty-self metaphor so enthusiastically. First, philosophical traditions that lend themselves to an empty-self perspective, such as the belief that people are born as a "blank slate" and subsequently shaped by the environment, have proven especially popular and influential in social psychology (Shkurko,

2019; von Hippel & Buss, 2017). Eventually, a focus on the influence of the social environment came to be incorporated into the very definition of social psychology, motivating researchers to advocate situational causes at the expense of dispositional causes and criticize disciplines that prioritized inter-situational consistency. Further, ascribing causal power to the situation is congenial to a progressive perspective, focused on combatting systematic inequalities and avoiding victim-blaming, that is overwhelmingly dominant among social psychologists (von Hippel & Buss, 2017)[3]. Finally, many social psychologists (and psychologists more broadly) have long wanted the discipline to be taken seriously as a "hard" or "real" science, leading to the prioritization of experimental methodologies over others (Masaryk & Stainton Rogers, 2024). This has encouraged researchers to design studies based on the questionable assumption that the self is empty and can therefore be readily manipulated. In turn, these experimental procedures often treat participants as if they were sterile chemicals, without a self that extended beyond the laboratory. Perhaps for these reasons, the empty-self metaphor has stubbornly persisted throughout the history of psychology, particularly in social psychology.

Despite its resilience, there are numerous reasons to doubt the veracity of the empty-self metaphor. Turning first to the issue of face validity, the empty-self metaphor is at odds with the subjective life experience of most people, who report durable beliefs about many topics (including themselves) and are reluctant to forfeit these beliefs. In a pre-replication crisis world, Nobel Prize winner Daniel Kahneman tellingly remarked that "when I describe priming studies to audiences, the reaction is often disbelief" (Kahneman, 2011). Similarly, forecasters overwhelmingly (and correctly) predicted that social priming research (e.g., Ackerman et al.,

[3] The empty-self metaphor can also support an authoritarian "law and order" conservative ideology, in which an intrusive state is needed to keep a reactive and volatile population under control. Given fewer than 2% of social psychologists are conservative (von Hippel & Buss, 2017), this right-wing version is seldom seen (with broken windows theory and moral panics over violence in rap music, television, and video games rare exceptions).

2010) would not replicate (Gordon et al., 2021). In fact, papers that garner a lot of media attention – as counterintuitive studies often do – are *less* likely to replicate (Youyou et al., 2023). The counterintuitive nature of a model is not sufficient reason to disbelieve it, but it should at least encourage skepticism and invite a high evidentiary burden.

More troublingly, the empty-self metaphor is at odds with large swaths of existing literature and theory. An entire field of psychology – Personality – is based on the well-founded assumption that the self is *not* empty, and generally employs a methodology based on people's capacity to reliably report their self-views. Personality measures do indeed appear to capture meaningful variance: they predict real-world behaviors and life outcomes in longitudinal studies (e.g., Iliescu et al., 2023), they show considerable stability over time (Anusic & Schimmack, 2016), and they concur with external observers' ratings of the self (Lee et al., 2024). Moreover, personality cannot be adequately explained by environmental factors alone, with 40-60% of the variance in the "Big Five" personality traits accounted for by genetics (Bouchard Jr & Loehlin, 2001). Relatedly, the empty-self metaphor is at odds with evolutionary theory, which argues that humans benefit from durable models of the self and others (Lewis & Buss, 2021; Sedikides et al., 2006). For example, there is ample evidence for stable sex differences in self-concept, which in turn can cause men and women to react differently to the exact same situation (Stake, 1992).

Of course, personality is not a *perfect* predictor of behavior (a ludicrously high standard) and there is evidence that personality varies over time and in how it is expressed between situations. Nevertheless, this does not undermine the validity of personality any more than seasonal fluctuations in weather undermine the concept of climate. Tellingly, there is considerable support for the predictive utility of personality traits. That is, the average effect size for personality traits versus situations is identical ($r = .21$; Fraley & Marks, 2007; Richard et al.,

2003). More compelling, just as personality psychology is the *most* successfully replicated subfield of psychology (Soto, 2019), social psychology is among the *least* replicable subfields (Open Science Collaboration, 2015; Youyou et al., 2023). The research literature therefore strongly suggests that pre-existing predispositions and mental models systematically shape behavior and govern the manner in which people experience the world. In fact, it is difficult to understand how anyone could effectively navigate the complexities of human interactions if unpredictable situational influences routinely overrode the self as the empty-self metaphor implies.

Even evidence that has been taken as proof of the power of situations to override moral principles or concrete sensory data has recently lost force. Consider the renowned obedience studies by Milgram, the poster child for the tendency for situational influences to overwhelm the self. The fact that the personalities of participants (acting as a "teacher") influenced their tendency to obey experimenter commands (Blass, 1991) is generally overlooked in accounts of this work. Furthermore, scrutiny of Milgram's lab notes indicates that he himself understood that his findings were not a simple demonstration of the capacity of the experimenter to bring teachers to heel. Instead, compliance hinged on the experimenter's facility in gaining the participant's allegiance and enlisting his agency in the successful conduct of the experiment. In light of this, Milgram even considered labeling his phenomenon "cooperation" rather than obedience (Haslam et al., 2015). Fittingly, a later conceptual replication found that conscientiousness and agreeableness, traits largely associated with positive and prosocial outcomes, predicted increased willingness to administer high-intensity electric shocks (Bègue et al., 2015). Similarly, contrary to the prevailing wisdom, participants in Asch's conformity studies rarely repudiated their own sensory data when performing a simple perceptual task (Deutsch &

Gerard, 1955). Instead, participants intentionally complied with the group consensus in an effort to let the errant (ostensibly misguided) experimental confederates save face (Swann & Jetten, 2017). And recent examination of features of the Stanford prison experiment have revealed that a quirk of the recruitment procedure (advertising the study as a psychological study of prison life) likely resulted in participants having elevated scores on authoritarianism (Carnahan & McFarland, 2007). This is noteworthy because authoritarianism would have predisposed guards to comply with Zimbardo's (the "prison warden") injunctions to bully prisoners and for prisoners to acquiesce to such bullying (Haslam et al., 2019).

Lastly, testimony to the limits of the power of the situation comes from evidence of people's powerful allegiance to their self-views. Findings inspired by self-verification theory suggest that firmly held self-views are so foundational to how humans interact with the world that they take active steps to maintain negative as well as positive self-views, even if it means soliciting evaluations that confirm negative conceptions of themselves and preferring relationships in which they are mistreated (Swann, 2012). Such self-verification strivings may help explain the failure of interventions designed to alter firmly held self-structures and their behavioral outcomes. For example, there is ample evidence that self-esteem remains relatively constant over the life span (Kuster & Orth, 2013; Orth & Robins, 2014), as do many other personal characteristics (e.g., gender identity, sexual preference). More generally, efforts to break habits frequently fail (Polivy & Herman, 2002), reallocating low-income families from high-crime to more prosperous communities has a negligible effect on criminal behavior (Harcourt & Ludwig, 2006), 30-55% of depression is treatment-resistant (McIntyre et al., 2023), and even brainwashing attempts have proved ineffective despite the fact that the change agents exerted complete control over the prisoners' environments (Schein, 1956). Given this, it is easy to

understand why more whimsical manipulations, such as plotting dots on a piece of paper (Pashler et al., 2012), would have a negligible influence on our models of the self and the world.

Of course, situational influences surely do influence cognition and behavior. At a fundamental level people are constantly reacting to their environment, even if they are unaware of its influence. For example, excessive heat may unknowingly contribute to feelings of irritability and aggression (Anderson et al., 2000)[4]. While acknowledging such influences, we contend that an interactionist account, in which a robust self is analyzed in the context of the situation, is the best predictor of behavior. That is, even if heat promotes irritability in general, an individual's pre-existing attitudes towards violence also play a large role in shaping their reactions to heat. Perhaps even more important, pre-existing attitudes and self-views can influence the type of situations people choose and how they respond to them. For example, people who are dispositionally aggressive or view themselves as such may be especially inclined to choose situations that elicit violence (Furr & Funder, 2021; Swann, 2012). The empty-self metaphor, however, would overlook the role of these pre-existing cognitive structures.

Empirically testing the association between the empty-self metaphor and replicability

Given the popularity of the empty-self metaphor, we examined whether hypotheses premised on the empty-self metaphor replicate less frequently. To test this possibility, we focused on two of the most prominent and highly-powered multi-lab replication studies in recent years; many labs 1 (Klein et al., 2014) and 2 (Klein et al., 2018). We coded 41 hypotheses as "empty-self" or not. These hypotheses spanned the history of psychology, from 1936 to 2014. We used three types of coders: AI (i.e., ChatGPT o3), a single informed coder (i.e., familiar with the

[4]This is surprisingly debatable, with embodied cognition research seemingly suggesting that warmth should promote prosociality rather than aggression, and recent empirical research failing to find *any* effect of ambient temperature on behavior and cognitions (Krause et al., 2023).

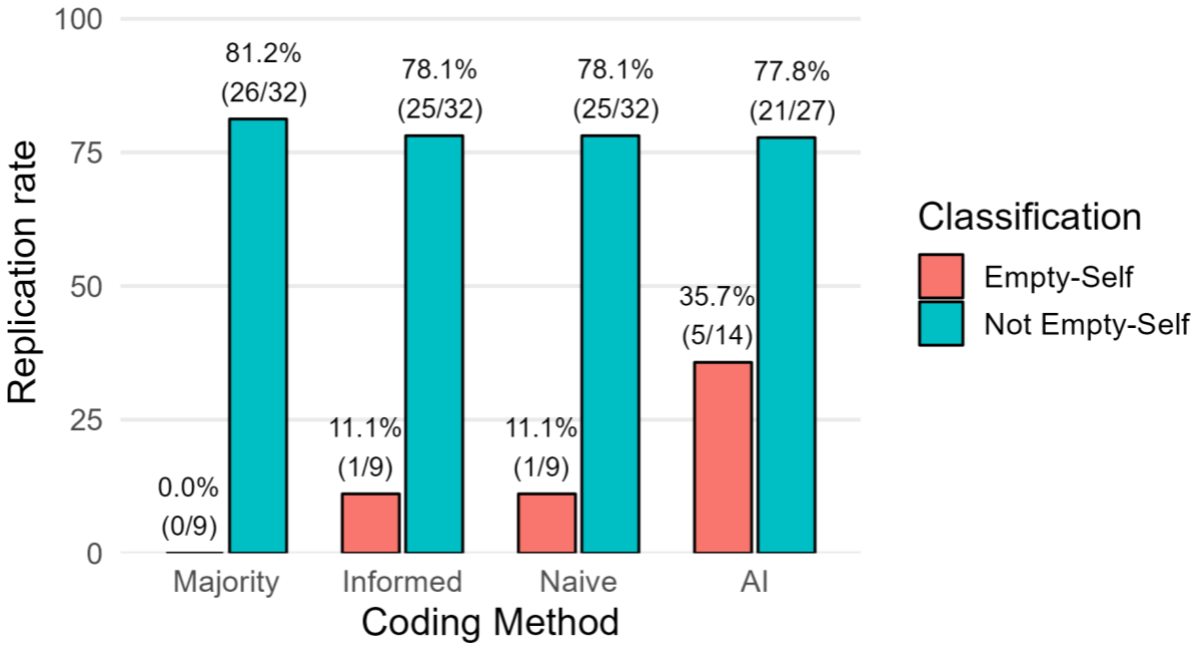
concept of the empty-self metaphor, but blind to which studies replicated), and three naïve coders (i.e., unfamiliar with the empty-self metaphor and blind to which studies replicated). Coders first learned that an empty-self hypothesis is one that assumes behavior is unexpectedly shaped by subtle environmental factors, often unconnected to the outcome variable, rather than by existing beliefs, dispositions, values etc., one would expect to be more important. This was differentiated from hypotheses that simply assumed the situation has an effect or did not include a personality measure, which does not necessarily imply an empty-self. Participants then coded 16 fictitious hypotheses as practice and were provided feedback. For example, the hypothesis "briefly showing a grey page background (vs. white) increases support for austerity policies" was classified as "empty-self" as it described a subtle environmental cue with no obvious link to the outcome measure. By contrast, "paying for a workshop out of pocket (vs. getting it free) increases attendance" was classified as *not empty-self* because the manipulation created a personal stake directly connected to the decision to attend. They then classified the focal 41 hypotheses.

The three naïve coders initially reported moderate agreement (AC1 = .56), with two coders reporting almost perfect agreement (AC1 = .86) and diverging from the third coder, who used the empty-self classification more frequently. The naïve coders then learned the codes for which they held the minority opinion and were allowed to reconsider their classifications. This procedure raised the agreement to substantial (AC1 = .67). Finally, the majority opinion was adopted for any remaining disagreements. The three coding types were in substantial agreement (AC1 = .76), with the two human coding strategies in near perfect agreement (AC1 = .85) and in substantial agreement with the AI (AC1 = .71).

Adopting the majority opinion across these three methods classified nine of the 41 hypotheses as “empty-self”. As can be seen in Figure 2, none of the hypotheses classified as empty-self replicated, versus 81.2% of those classified as not empty-self. Follow-up analyses using Fisher’s Exact Test suggested a significant association between a hypothesis being classified as empty-self and non-replication ($p < .001$), with a strong effect size (Cramér’s V = .56). Additional analyses using each of the coding methods individually yielded similar results, with AI coding having the weakest effect size (Cramér’s V = .36). Robustness tests, in which analyses were restricted to hypotheses with unanimous classification across the three coding methods or with “weak replications” recoded as non-replications, had a negligible effect on results. All code, data, and instructions are shared on the Open Science Framework (https://osf.io/y9pwn/?view_only=246da9b0820d494da8107bb4aed9f456). These results provide preliminary evidence that hypotheses based on the empty-self metaphor may be more difficult to replicate.

Figure 2

Replication rate of hypotheses adopting the empty-self vs. not empty-self metaphor in many labs 1 and 2 studies under different coding strategies



The empty-self metaphor today

Were the empty-self metaphor merely a relic of a prior century that has faded into oblivion, it would hardly deserve mention. But what is commonly overlooked is that the empty-self metaphor can still be found in certain quarters of contemporary social psychology. For example, a relatively recent review of moral psychology claimed that "seemingly unimportant or irrelevant situational features can have far-reaching implications for real-life moral decisions" (Ellemers et al., 2019; pg. 357). Striking a decidedly neo-situationist tone, the authors lamented that "it is difficult to understand why so many researchers still rely on measures that capture individual differences or general tendencies and assume these have predictive value across situations" (Ellemers et al., 2019; pg. 356). Elsewhere, Gino and Ariely (2016; pg. 336) argued that "situational forces are so strong that they make individual choice all but irrelevant". The

boldness of this claim was widely celebrated, at least until much of the work underpinning it was found to be fraudulent (O'Grady, 2023).

We speculate that the continued influence of the empty-self metaphor, and the shaky theoretical foundations it has promulgated, helps to explain the replication crisis we are presently contending with. Consider a few theories that have occupied center stage in social psychology until quite recently. Social priming, which contends that imperceptible cues can have a large effect on human behavior, was once a hugely acclaimed field of social psychology. Similarly, the theory of embodied cognition, which was introduced under the priming umbrella, suggested that the environment can have a "metaphorical power" on human cognitions. Most famously, Williams and Bargh claimed to much fanfare that holding a warm cup of coffee caused participants to judge others as being "warmer", seemingly internalizing the metaphorical power of the situation (2008). It seemed unbelievable that such an innocuous environmental cue could have a measurable impact on behavior – one that could override a person's natural tendency to view others as agreeable or not – and subsequent years have proven this skepticism to be well warranted. A host of priming effects, such as claims that warmth from a shower or bath can compensate for social warmth (Bargh & Shalev, 2012), have failed to replicate (Donnellan et al., 2015). In fact, out of 70 close replications of priming studies, 94% reported smaller effect sizes than the original, and only 17% found a significant effect in the expected direction (Mac Giolla et al., 2024). The related formulation of "power posing", in which adopting an expansive posture was purported to bolster confidence among those that lacked it, proved extremely popular with the general public but has had a mixed replication track record (Körner et al., 2022). This widespread inability to replicate original findings has largely dimmed interest in social priming

(Chivers, 2019), with even Kahneman pronouncing the field "effectively dead" (Kahneman, 2022).

Aspects of other empty-self formulations have likewise triggered incredulity of late. The classic social identity approach, for instance, contended that when people enter groups a "depersonalization" process causes them to focus on a relevant social self (e.g. "I am an American") at the cost of their personal selves (e.g., introverted, intelligent). This reasoning led to the conclusion that enduring personal characteristics become irrelevant in group situations (Turner, 1985). In subsequent decades, a host of theorists and researchers (Abrams, 1994; Greenaway et al., 2015; Huddy, 2001, 2002; Pickett et al., 2002; Reid & Deaux, 1996; Simon, 2004; Spears, 2001) questioned the depersonalization argument, suggesting that stable conceptions of the self can play a key role in group contexts. For example, challenging the chronic self-views of people who were strongly aligned with a group increased their willingness to fight and die for the group (experiments 1 and 2; Swann et al., 2009). Later studies indicated that feeling understood at a personal level increased the student's affinity for, and willingness to make extreme sacrifices for, other group members (Gómez et al., 2024). Similarly, when incels – people who belligerently self-identify as being involuntarily celibate – felt understood by other incels, they were more inclined to engage in the pro-group behavior of endorsing harassment of women (whom they perceived as outgroup members; Rousis et al., 2023). Together, this research suggests that the self plays an important role in situations in which group membership is highly salient.

More broadly, we can see the empty-self metaphor persist across the various subfields of psychology. In clinical psychology some have decried the popularity of a "trauma-centric" view of psychopathology in which childhood trauma is described as the "overwhelming major driver

of psychopathology in Western civilization" (pg. 122; Ross & Pam, 2005). Critics have lambasted this view as reflecting a radical environmentalism in which dispositional and genetic contributors to psychopathology and resilience are overlooked (Lilienfeld, 2010). Elsewhere, positive psychologists recommend positive self-talk (e.g., reciting "I am a lovable person") to improve wellbeing, even though it tends to make the people who really need it – those with low self-esteem – feel worse (Wood et al., 2009). Behavioral economists have popularized the use of nudges typically without regard for enduring individual differences that may influence their effectiveness (Peer et al., 2020). Indeed, there is limited meta-analytic evidence that "universal" nudges work when accounting for publication bias (Maier et al., 2022), with a recent well-powered, pre-registered study finding evidence for a null effect for nudging honesty (Dimant et al., 2020). Similarly, although some political psychologists have suggested that subtle linguistic cues, such as describing someone as a "voter" rather than "one who votes", activated a person's social identity and raised turnout by an incredible 11-14 points (Bryan et al., 2011), subsequent studies failed to replicate these results (Gerber et al., 2018). Finally, a recent critique of peace psychology, tasked with developing psychological strategies for promoting peace and avoiding war, forcefully argued that the field is dominated by a *naïve pacifism* that treats violence as a reaction to a provocative or threatening environment (Adam-Troian et al., 2024). This approach disregards the varying (often nefarious) motives underlying violent behaviors, leading to the erroneous assumption that nations "sleepwalk" into war and discounting how individuals can create conflict to advance their own strategic ambitions.

Although the foregoing evidence challenges some of the more extreme ideas based on the empty-self metaphor, it does not impugn the motives of the original theorists or researchers. Unfortunately, the exciting, counterintuitive results based on the empty-self metaphor – almost

guaranteeing publication in a top journal – have proven to be so enticing that some have pursued outright fraud to produce them. Work by discredited social psychologist Diedrik Stapel, who has candidly admitted to having invented data (Callaway, 2011), examined how seemingly innocuous parts of the environment can have a huge impact on behavior. For example, one of his most famous (now retracted) articles supposedly found that being exposed to litter or an abandoned bicycle could somehow promote discrimination (Stapel & Lindenberg, 2011). Stapel argued that "the message for policy-makers was clear" (Stapel & Lindenberg, 2011; p. 253); interventions as simple as cleaning up are helpful in the fight against racism. Likewise, retracted work by famed psychologists Francesca Gino and Dan Ariely suggested that signing a document at the beginning, rather than at the end of a self-report form, greatly reduced dishonesty (Shu et al., 2012). Both of these highly cited articles implied that minor interventions – cleaning streets or changing the location of a signature – could have a substantial impact on behavior, overriding existing inclinations towards discrimination or lying. Unfortunately, both were based on fraudulent data, and have thus cast further doubt on the empty-self metaphor.

Skeptics could point out that current iterations of the empty-self metaphor are less explicit than previous iterations. Most contemporary social psychologists will at least pay lip service to an enduring self, even if its role is often minimized in their actual work. For example, self-categorization theory (of the social identity approach) does acknowledge the self with its concept of perceiver readiness, stipulating that people may be more favorably disposed toward certain groups based on their prior experience (Turner et al., 1994). However, this is an underappreciated aspect of the theory, warranting only a brief mention in a widely-cited history of the social identity approach (Hornsey, 2008). Social priming theorists have argued that successful priming requires that people already possess the relevant stereotype in the first place,

meaning that the self cannot be *entirely* empty. In fact, Bargh suggested that a failure to replicate a study in which participants walked slower when exposed to elderly stereotypes (Doyen et al., 2012) may have been because the replicators did not confirm that participants held the primed stereotype (Bargh, 2012), although *he* did not do this either in his original study (Bargh et al., 1996). Even granting this, the study still does not allow much room for a coherent self; assuming the undergraduate sample did not consider themselves to be elderly, we are left with the unlikely proposition that people readily adopt whatever stereotypes are situationally salient, even when such stereotypes clash with their extant self-views. So, while we recognize that most modern social psychologists may superficially acknowledge the self, the empty-self metaphor often looms large.

Indeed, even in this implicit form, the empty-self metaphor has the potential to promote faulty beliefs and questionable recommendations. For instance, a theme in the embodied cognition literature is that moral cleansing can be metaphorically attained via actual cleansing. One well-cited article implies that, although "physical cleanliness has many medical benefits", this must be weighed against the potential dangers of cleanliness enhancing moral self-perceptions that license harsher moral judgements on issues such as abortion and pornography (pg. 859; Zhong et al., 2010). Putting aside concerns regarding the replicability of this line of research (e.g., "Lady Macbeth effect"; Earp et al., 2014; Siev et al., 2018), it would surely be worrying if research based on such a questionable premise caused people to reconsider washing their hands. Clearly, a field predicated upon sensational, counterintuitive results, such as those that have sprung from the empty-self metaphor, can foster dubious outcomes.

Future Directions

We contend that the empty-self metaphor has fostered, and continues to foster, an incomplete and misleading portrait of human beings. This portrait, in turn, encourages research that overlooks the role of an enduring self in shaping people's reactions to the world around them and instead offers simple solutions to complex problems. The appeal of sensational and counterintuitive findings, in combination with questionable research practices that can make practically any desired result statistically significant (Simmons et al., 2011), creates a large incentive to report results that support the empty-self metaphor. It is no wonder then that the findings inspired by this metaphor have proven challenging to replicate.

The good news is that the antidote to this problem is already available. Decades ago, Kurt Lewin (Lewin, 1946) challenged Watson's (Watson, 1925) boast that he could successfully shape the behavior of individuals while disregarding their individual qualities. Lewin rejected this behavioristic credo, arguing instead for a more expansive view that treated persons and situations as equal partners in shaping behavior. Although Lewin's influential *Behavior = f (Person, Situation)* equation was overshadowed by situationism in subsequent years, it has regained traction in the last several decades (Bond, 2013; Funder, 2006; Swann & Seyle, 2005). Others have made relevant modifications to this principle; for instance, in the same way people only acquire viruses that match receptors they possess, people are only influenced by situations they are sufficiently "tuned" to (Swann & Seyle, 2005). With this refined formulation we can surmise, for example, that primes with no relevance to the self are unlikely to influence behavior. We propose that this new trend could be further strengthened by the explicit recognition that the self is not an empty vessel but an active force that guides people's responses to situations.

Taking this insight seriously has important methodological implications. Broadly speaking, models that consider an interaction between the situation and the self will tend to

account for more variance than those that consider the situation alone (Kuper et al., 2024). The prevalence of comprehensive personality taxonomies (e.g., the Big 5; McCrae, 1999; Soto & John, 2017), along with recent efforts to categorize the psychological characteristics of situations (Rauthmann et al., 2020), have opened the door for a more systematic examination of person x situation interactions. Nevertheless, difficulties in precisely pinpointing the specific elements of personality or the situation that underlie a given interactive effect means that further research is needed to better refine our understanding of the processes underlying this interplay (Kuper et al., 2024).

Moreover, it is worth reconsidering whether experiments are a realistic option for every research question. Experimental manipulations have historically been treated as the "gold standard" methodology for examining social psychological phenomenon, while qualitative, quasi-experimental, and non-experimental approaches have been unfortunately often dismissed (Diener et al., 2022; Swann & Jetten, 2017). While there are clear reasons for this preference – the most obvious being that experimentation can establish causal relationships – it may be time to acknowledge that many aspects of the self cannot be easily or reliably manipulated. For example, poor and incoherent experimental manipulations have been cited as a key reason that research on ego depletion has proven difficult to replicate (Eronen & Bringmann, 2021). Likewise, the often brief and superficial nature of some growth mindset interventions, such as reading a short article, may explain failures to produce replicable effects (Brez et al., 2020). This reliance on subtle manipulations could go a long way towards explaining why studies using experimental methods are less replicable than those using non-experimental methods (Youyou et al., 2023). Abandoning the empty-self metaphor and acknowledging that participants are not

vacuous blank canvases should encourage social psychologists to be realistic about what can be experimentally manipulated, and to be more willing to employ other methodologies.

Even the way experiments are commonly analyzed overlooks the self. Often a t-test or Analysis of Variance (ANOVA) is performed to compare the mean scores of an experimental and control condition. However, doing so suggests that all participants in a condition are alike and are equally affected by the manipulation. This is rarely the case. More broadly, describing behavior in terms of groups (i.e., conditions), when participants are actually acting as individuals, may be statistically convenient but misleading. For these reasons, some have argued that researchers should indicate how many participants were affected by a manipulation (e.g., 10% of participants in the experimental condition increased their score on the dependent variable), thereby reflecting the diversity of responses from the people that comprise a study's sample (Billig, 2013). Even simple data visualizations that illustrate the range of responses, like a violin or density plot, would help demonstrate individual differences to standardized environmental manipulations. Doing so will more accurately portray the "power", or lack thereof, of the situation.

In fact, jettisoning the empty-self metaphor may necessitate a change in the very way we write about social psychology. In an attempt to mirror the "hard" sciences, experimental social psychologists tend to passively describe their results in terms of technical, impersonal processes, like one might pasteurization or corrosion, rather than actual, messy human behavior (Billig, 2013). For example, one might say that variable X (e.g., ego-depletion, priming) had an effect on variable Y (e.g., self-control, goal-pursuit), without ever making reference to an actual person. To be fair, this issue is discipline-wide and the authors of this article are certainly guilty of writing in this way too. However, the problem is exacerbated by a belief that the self is empty, which in

turn increases the ease with which the self can be obscured. Various remedies have already been prescribed: writing in the active voice, a willingness to use verbs, and preferring simple over technical terms (Billig, 2013). But above all, it would behoove social psychologists to remember that real *people* lie at the heart of the field, and to write accordingly.

It is worth pointing to examples in which theories discussed in the present article have considered the prospect of an enduring self, thereby improving the reliability of their findings. For example, while the minimal groups paradigm in social identity theory – in which people demonstrate bias towards an ingroup defined by arbitrary or trivial selection criteria (e.g., the toss of a coin) – has generally proven to be a robust phenomenon, researchers have occasionally failed to replicate it (Kerr et al., 2018). To better understand these failed replications, Kerr et al. (2018) systematically examined differences in experimental design and found that, among other factors, the cultural background of the participants influenced results (with the egalitarian-minded Australian participants less biased towards the ingroup than Americans). By considering aspects of an enduring self that extend beyond the lab – in this case, culturally acquired and internalized values – the researchers were able to identify boundary conditions and strengthen the replicability of the paradigm. Likewise, more recent "second generation" nudges are increasingly personalized to the target (Mills, 2022). Recent research found that personalized reminders to make healthier food purchases (e.g., by matching nudges to whether the person was motivated by health or price, or preferred information presented visually or verbally) were more effective in promoting positive behavior than generic reminders, and had a small "spillover" effect in which participants continued to make healthier choices (de Vries et al., 2025). Likewise, others have argued that behavior-change interventions are most effective when they account for a person's enduring characteristics (Rebele et al., 2021). For instance, tailoring password-strength

nudges to people's dispositional decision-making styles increased their effectiveness by up to four times (Peer et al., 2020). Research like this highlights the usefulness of considering the influence of an enduring self in experimental designs and interventions.

In general, we should reiterate that we have enormous respect for the contributions of many of the theories criticized in this article. This is particularly true for foundational theories, such as social identity theory and cognitive dissonance, which have had a profound impact on the way we understand social phenomena. Even the most sensational claims from the social priming literature arose from more modest and intuitive theories that are surely valid (e.g., semantic priming). Rather than seeking to criticize the field writ large, our aim is to highlight that these theories could all be improved by replacing the empty-self assumption with a framework that allows for an enduring self. There has already been some progress towards this goal. For instance, social identity researchers have increasingly explored chronic individual differences in group identification, helping to address concerns that the theory ignores enduring features of the self (e.g., Huddy, 2002)[5]. We articulate specific suggestions in Table 1.

Table 1

Prototypical social psychology approaches and the empty-self metaphor

Theory	Empty-Self Assumption	Enduring-Self Assumption
Social priming and embodied cognition	A large number of subtle environmental cues determine behavior, regardless of personal relevance	A select number of salient environmental cues influence behavior, depending on personal relevance
Social identity theory	The social self determines behavior when group membership is salient	The personal and social selves can both influence behavior when group membership is salient
Nudge Theory	Impersonal nudges guide behavior without regard to a unique, enduring self	Personalized nudges guide behavior by targeting relevant elements of a unique, enduring self

[5] That said, the social identity approach emphasis on a tension between personal and social self-views could complicate efforts to embrace a conventional person x situation interactionist approach. Moreover, social identity theorists still tend to emphasize the instability of the self (e.g., Cruwys et al., 2025).

Self-perception theory	Self-views are epi-phenomenal constructions based on observations of one's own behavior and the situations in which it occurs	Self-views are enduring and influential determinants of thought and action that can be strengthened or weakened via observing one's behavior
Situationism	Situations regulate behavior regardless of qualities of persons	Characteristics of persons and situations determine behavior interactively or additively

Abandoning the empty-self metaphor may not only strengthen the formulations it has compromised, but also lay the groundwork for a broader reproachment between social and personality psychology. Let us be clear: we are not advocating that either social or personality psychology be subordinated to the other; rather, we are urging greater openness to the many ways in which each subdiscipline can learn from the other. This could benefit both subdisciplines. We have already noted that recognizing the importance of the self could benefit social psychology by, for instance, bolstering the replicability of effects. But personality psychology could also benefit from social psychology's focus on the psychological mechanisms underlying behavior. That is, personality psychologists – especially those focusing on traits – have occasionally struggled to move beyond demonstrating the construct validity of the elegant measures they have devised. Attention to the processes that underlie the individual differences they have identified – using theory and methods devised by social psychologists – could enrich these formulations. More generally, explicitly recognizing that both subdisciplines bring important insights to the table cannot but help foster meaningful integration and cross-fertilization. To this end, joint conferences of the two subdisciplines and normalization of hybrid, social-personality programs (rather than separate social and personality programs) might encourage more constructive inter-play between the two.

Finally, it is worth returning to what makes the empty-self metaphor so popular: (1) the appeal of a minor intervention that produces major effects, and (2) the entertainment value of its

counterintuitive or surprising implications. On the first point, the adage “extraordinary claims require extraordinary evidence” serves us well. If the results of a study appear incredible then they should be treated with caution until they are replicated in a pre-registered study by disinterested researchers, given that researchers with financial incentives tied to producing a positive result are much more likely to find stronger and significant effects (Mac Giolla et al., 2024; Macnamara & Burgoyne, 2023). On the second point we would all benefit from recalibrating our attitudes towards counterintuitive results in psychology; such findings should first prompt skepticism, not celebration. That is not to say that our intuitions are always correct, but that the accuracy of prediction markets that forecast study replicability suggests they are a useful guide (e.g., Gordon et al., 2021). We are essentially encouraging a Bayesian epistemology, in which the degree of attention paid to a result is proportional to the a-priori likelihood of it being true, and astounding findings are treated with skepticism until the supporting evidence is sufficiently overwhelming.

Conclusion

The metaphor of the empty-self can be seen throughout the history of psychology, from behaviorism until the present day. Unfortunately, it has little basis in reality and may have contributed to the replication crisis that has rocked psychology. We contend that the self is far from empty, and that both the person and the situation — either alone or interactively — must be considered by psychologists. By switching to a metaphor of the self as an enduring and active agent, rather than an inconsequential empty vessel, social psychologists will be better positioned to generate findings that will allow the field to reach its full potential.

Reference List

Abelson, R. P. (1983). Whatever became of consistency theory? *Personality and Social Psychology Bulletin*, *9*(1), 37-64. https://doi.org/10.1177/0146167283091006

Abrams, D. (1994). Social self-regulation. *Personality and Social Psychology Bulletin*, *20*(5), 473-483. https://doi.org/10.1177/0146167294205004

Ackerman, J. M., Nocera, C. C., & Bargh, J. A. (2010). Incidental haptic sensations influence social judgments and decision. *Science*, *328*(5986), 1712-1715. https://doi.org/10.1126/science.1189993

Adam-Troian, J., Arciszewski, T., Belanger, J., & Bonetto, E. (2024). Pacifist or peace psychology? Overcoming naïve pacifism to foster a true science of peace-making. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/naj2x

Allport, F. H. (1924). *Social Psychology*. Houghton Mifflin. https://books.google.com.hk/books?id=OcwEAQAAIAAJ

Allport, G. W. (1954). *The Nature of Prejudice*. Addison-Wesley.

Anderson, C. A., Anderson, K. B., Dorr, N., DeNeve, K. M., & Flanagan, M. (2000). Temperature and aggression. In *Advances in Experimental Social Psychology* (Vol. 32, pp. 63-133). Academic Press. https://doi.org/10.1016/S0065-2601(00)80004-0

Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology*, *110*(5), 766-781. https://doi.org/10.1037/pspp0000066

Aronson, E. (1968). Dissonance theory: Progress and problems. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Rand McNally.

Aronson, E. (2019). Dissonance, hypocrisy, and the self-concept. In *Cognitive dissonance: Reexamining a pivotal theory in psychology, 2nd ed.* (pp. 141-157). American Psychological Association. https://doi.org/10.1037/0000135-007

Aronson, E., & Carlsmith, J. M. (1962). Performance expectancy as a determinant of actual performance. *The Journal of Abnormal and Social Psychology*, *65*(3), 178-182. https://doi.org/10.1037/h0042291

Baggini, J. (2011). *The Ego Trick*. Granta. https://books.google.com.hk/books?id=4RqqdXfp6fgC

Bargh, J. A. (2012). Nothing in their heads. *The Natural Unconscious*.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230-244. https://doi.org/10.1037/0022-3514.71.2.230

Bargh, J. A., & Shalev, I. (2012). The substitutability of physical and social warmth in daily life. *Emotion*, *12*(1), 154-162. https://doi.org/10.1037/a0023527

Baumeister, R. F., & Hutton, D. G. (1987). Self-presentation theory: Self-construction and audience pleasing. In B. Mullen & G. R. Goethals (Eds.), *Theories of Group Behavior* (pp. 71-87). Springer New York. https://doi.org/10.1007/978-1-4612-4634-3_4

Bègue, L., Beauvois, J. L., Courbet, D., Oberlé, D., Lepage, J., & Duke, A. A. (2015). Personality predicts obedience in a Milgram paradigm. *J Pers*, *83*(3), 299-306. https://doi.org/10.1111/jopy.12104

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6, pp. 1-62). Academic Press. https://doi.org/10.1016/S0065-2601(08)60024-6

Billig, M. (2013). *Learn to Write Badly: How to Succeed in the Social Sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9781139208833

Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, *60*(3), 398-413. https://doi.org/10.1037/0022-3514.60.3.398

Bond, M. H. (2013). Refining Lewin's formula: A general model for explaining situational influence on individual social behavior. *Asian Journal of Social Psychology*, *16*(1), 1-15. https://doi.org/10.1111/ajsp.12012

Bouchard Jr, T. J., & Loehlin, J. C. (2001). Genes, evolution, and personality. *Behavior Genetics*, *31*(3), 243-273. https://doi.org/10.1023/A:1012294324713

Brez, C., Hampton, E. M., Behrendt, L., Brown, L., & Powers, J. (2020). Failure to replicate: Testing a growth mindset intervention for college student success. *Basic and Applied Social Psychology*, *42*(6), 460-468. https://doi.org/10.1080/01973533.2020.1806845

Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences*, *108*(31), 12653-12656. https://doi.org/10.1073/pnas.1103343108

Callaway, E. (2011). Report finds massive fraud at Dutch universities. *Nature*, *479*(7371), 15-15. https://doi.org/10.1038/479015a

Carnahan, T., & McFarland, S. (2007). Revisiting the Stanford prison experiment: Could participant self-selection have led to the cruelty? *Personality & Social Psychology Bulletin*, *33*(5), 603-614. https://doi.org/10.1177/0146167206292689

Chabris, C. F., Heck, P. R., Mandart, J., Benjamin, D. J., & Simons, D. J. (2019). No evidence that experiencing physical warmth promotes interpersonal warmth: Two failures to replicate Williams and Bargh (2008). *Social Psychology*, *50*(2), 127-132. https://doi.org/10.1027/1864-9335/a000361

Chaiken, S., & Baldwin, M. W. (1981). Affective-cognitive consistency and the effect of salient behavioral information on the self-perception of attitudes. *Journal of Personality and Social Psychology*, *41*(1), 1-12. https://doi.org/10.1037/0022-3514.41.1.1

Chivers, T. (2019). What's next for psychology's embattled field of social priming. *Nature*, *576*, 200-202. https://doi.org/10.1038/d41586-019-03755-2

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

Cruwys, T., Haslam, S. A., & Skorich, D. P. (2025). Disorders of self-categorization: How and why a healthy social self-system is the cornerstone of mental health. *Psychological Review*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/rev0000566

de Vries, R., Bol, N., & van der Laan, N. (2025). "Just-in-time" but a bit delayed: Personalizing digital nudges for healthier online food choices. *Appetite*, *206*, 107852. https://doi.org/10.1016/j.appet.2025.107852

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629-636. https://doi.org/10.1037/h0046408

Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, *17*(4), 1101-1119. https://doi.org/10.1177/17456916211037670

Dimant, E., van Kleef, G. A., & Shalvi, S. (2020). Requiem for a Nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization*, *172*, 247-266. https://doi.org/https://doi.org/10.1016/j.jebo.2020.02.015

Donnellan, M. B., Lucas, R. E., & Cesario, J. (2015). On the association between loneliness and bathing habits: Nine replications of Bargh and Shalev (2012) Study 1. *Emotion*, *15*(1), 109-119. https://doi.org/10.1037/a0036079

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), e29081. https://doi.org/10.1371/journal.pone.0029081

Earp, B. D., C., E. J. A., N., M. E., & and Hamlin, J. K. (2014). Out, damned spot: Can the "Macbeth Effect" be replicated? *Basic and Applied Social Psychology*, *36*(1), 91-98. https://doi.org/10.1080/01973533.2013.856792

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: A Review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, *23*(4), 332-366. https://doi.org/10.1177/1088868318811759

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

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Fraley, R. C., & Marks, M. J. (2007). The null hypothesis significance-testing debate and its implications for personality research. In *Handbook of research methods in personality psychology.* (pp. 149-169). The Guilford Press.

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40*(1), 21-34. https://doi.org/10.1016/j.jrp.2005.08.003

Furr, R. M., & Funder, D. C. (2021). Persons, situations, and person–situation interactions. In *Handbook of personality: Theory and research* (Vol. 4, pp. 667-685). The Guilford Press.

Gerber, A., Huber, G., & Fang, A. (2018). Do subtle linguistic interventions priming a social identity as a voter have outsized effects on voter turnout? Evidence from a new replication experiment. *Political Psychology*, *39*(4), 925-938. https://doi.org/https://doi.org/10.1111/pops.12446

Gino, F., & Ariely, D. (2016). Dishonesty explained: What leads moral people to act immorally. In A. G. Miller (Ed.), *The Social Psychology of Good and Evil* (2 ed.). Guilford Press.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Doubleday.

Gómez, Á., Vázquez, A., Alba, B., Blanco, L., Chinchilla, J., Chiclana, S., & Swann, W. B. (2024). Feeling understood fosters identity fusion. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000464

Gordon, M., Viganola, D., Dreber, A., Johannesson, M., & Pfeiffer, T. (2021). Predicting replicability-Analysis of survey and prediction market data from large-scale forecasting projects. *PLoS ONE*, *16*(4), e0248780. https://doi.org/10.1371/journal.pone.0248780

Greenaway, K. H., Haslam, S. A., Cruwys, T., Branscombe, N. R., Ysseldyk, R., & Heldreth, C. (2015). From "we" to "me": Group identification enhances perceived personal control with consequences for health and well-being. *Journal of Personality and Social Psychology*, *109*(1), 53-74. https://doi.org/10.1037/pspi0000019

Harcourt, B. E., & Ludwig, J. (2006). Broken windows: New evidence from New York City and a five-city social experiment. *The University of Chicago Law Review*, *73*(1), 271-320. http://www.jstor.org/stable/4495553

Haslam, S. A., Reicher, S. D., Millard, K., & McDonald, R. (2015). 'Happy to have been of service': The Yale archive as a window into the engaged followership of participants in Milgram's 'obedience' experiments. *British Journal of Social Psychology*, *54*(1), 55-83. https://doi.org/10.1111/bjso.12074

Haslam, S. A., Reicher, S. D., & Van Bavel, J. J. (2019). Rethinking the nature of cruelty: The role of identity leadership in the Stanford Prison Experiment. *American Psychologist*, *74*(7), 809-822. https://doi.org/10.1037/amp0000443

Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, *2*(1), 204-222. https://doi.org/10.1111/j.1751-9004.2007.00066.x

Huddy, L. (2001). From social to political identity: A critical examination of social identity theory. *Political Psychology*, *22*(1), 127-156. https://doi.org/https://doi.org/10.1111/0162-895X.00230

Huddy, L. (2002). Context and meaning in social identity theory: A response to Oakes. *Political Psychology*, *23*(4), 825-838. http://www.jstor.org/stable/3792369

Ichheiser, G. (1943). Misinterpretations of personality in everyday life and the psychologist's frame of reference. *Journal of Personality*, *12*(2), 145-160. https://doi.org/10.1111/j.1467-6494.1943.tb01953.x

Iliescu, D., Ion, A., Ilie, A., Ispas, D., & Butucescu, A. (2023). The incremental validity of personality over time in predicting job performance, voluntary turnover, and career success in high-stakes contexts- A longitudinal study. *Personality and Individual Differences*, *213*, 112288. https://doi.org/https://doi.org/10.1016/j.paid.2023.112288

Jones, E. E. (1964). *Ingratiation: A Social Psychologist Analysis.* Appleton-Century-Croft.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kahneman, D. (2022). Adversarial collaboration: An EDGE lecture by Daniel Kahneman https://www.edge.org/adversarial-collaboration-daniel-kahneman

Katz, D. (1979). Obituary: Floyd H. Allport (1890-1978). *American Psychologist*, *34*(4), 351-353. https://doi.org/10.1037/h0078276

Kerr, N. L., Ao, X., Hogg, M. A., & Zhang, J. (2018). Addressing replicability concerns via adversarial collaboration: Discovering hidden moderators of the minimal intergroup discrimination effect. *Journal of Experimental Social Psychology*, *78*, 66-76. https://doi.org/https://doi.org/10.1016/j.jesp.2018.05.001

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M.,…Nosek, B. A. (2014). Investigating variation in replicability: A "many labs"

replication project. *Social Psychology*, *45*(3), 142-152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R.,…et al. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443-490. https://doi.org/10.1177/2515245918810225

Körner, R., Röseler, L., Schütz, A., & Bushman, B. J. (2022). Dominance and prestige: Meta-analytic review of experimentally induced body position effects on behavioral, self-report, and physiological dependent variables. *Psychological Bulletin*, *148*(1-2), 67-85. https://doi.org/10.1037/bul0000356

Krause, J. S., Brandt, G., Schmidt, U., & Schunk, D. (2023). Don't sweat it: Ambient temperature does not affect social behavior and perception. *Journal of Economic Psychology*, *99*, 102657. https://doi.org/10.1016/j.joep.2023.102657

Kuper, N., von Garrel, A. S., Wiernik, B. M., Phan, L. V., Modersitzki, N., & Rauthmann, J. F. (2024). Distinguishing four types of Person × Situation interactions: An integrative framework and empirical examination. *Journal of Personality and Social Psychology*, *126*(2), 282-311. https://doi.org/10.1037/pspp0000473

Kuster, F., & Orth, U. (2013). The long-term stability of self-esteem: Its time-dependent decay and nonzero asymptote. *Personality and Social Psychology Bulletin*, *39*(5), 677-690. https://doi.org/10.1177/0146167213480189

Lee, K., de Vries, R. E., & Ashton, M. C. (2024). Self/observer agreement in personality assessment by observers' relationship types. *Journal of Research in Personality*, *113*, 104529. https://doi.org/https://doi.org/10.1016/j.jrp.2024.104529

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, *2, 4*, 34-46. https://doi.org/10.1111/j.1540-4560.1946.tb02295.x

Lewis, D. M. G., & Buss, D. M. (2021). The evolution of human personality. In *Handbook of Personality: Theory and Research* (Vol. 4, pp. 3-34). The Guilford Press.

Lilienfeld, S. O. (2010). Can psychology become a science? *Personality and Individual Differences*, *49*(4), 281-288. https://doi.org/https://doi.org/10.1016/j.paid.2010.01.024

Mac Giolla, E., Karlsson, S., Neequaye, D. A., & Bergquist, M. (2024). Evaluating the replicability of social priming studies. *Meta-Psychology*, *8*. https://doi.org/10.15626/MP.2022.3308

Macnamara, B. N., & Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices. *Psychological Bulletin*, *149*(3-4), 133-173. https://doi.org/10.1037/bul0000352

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J. L., & Wagenmakers, E. J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences*, *119*(31), e2200300119. https://doi.org/10.1073/pnas.2200300119

Masaryk, R., & Stainton Rogers, W. (2024). The time has come for psychology to stop treating qualitative data as an embarrassing secret. *Social and Personality Psychology Compass*, *18*(2), e12938. https://doi.org/https://doi.org/10.1111/spc3.12938

McCrae, R., & Costa, P. (1999). The five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2 ed., pp. 139-153). Guilford.

McIntyre, R. S., Alsuwaidan, M., Baune, B. T., Berk, M., Demyttenaere, K., Goldberg, J. F., Gorwood, P., Ho, R., Kasper, S., Kennedy, S. H., Ly-Uson, J., Mansur, R. B., McAllister-Williams, R. H., Murrough, J. W., Nemeroff, C. B., Nierenberg, A. A., Rosenblat, J. D., Sanacora, G., Schatzberg, A. F.,…Maj, M. (2023). Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. *World Psychiatry*, *22*(3), 394-412. https://doi.org/10.1002/wps.21120

Mills, S. (2022). Personalized nudging. *Behavioural Public Policy*, *6*(1), 150-159. https://doi.org/10.1017/bpp.2020.7

Mischel, W. (1968). *Personality and Assessment*. Wiley. https://books.google.com.au/books?id=r999AAAAMAAJ

Mischel, W. (2009). From personality and assessment (1968) to personality science, 2009. *Journal of Research in Personality*, *43*(2), 282-290. https://doi.org/https://doi.org/10.1016/j.jrp.2008.12.037

O'Grady, C. (2023). Harvard behavioral scientist faces research fraud allegations. *Science*. https://doi.org/10.1126/science.adj3539

Orth, U., & Robins, R. W. (2014). The development of self-esteem. *Current Directions in Psychological Science*, *23*(5), 381-387. https://doi.org/10.1177/0963721414547414

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS ONE*, *7*(8), e42510. https://doi.org/10.1371/journal.pone.0042510

Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., & Frik, A. (2020). Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior*, *109*, 106347. https://doi.org/https://doi.org/10.1016/j.chb.2020.106347

Pickett, C. L., Silver, M. D., & Brewer, M. B. (2002). The impact of assimilation and differentiation needs on perceived group importance and judgments of ingroup size. *Personality and Social Psychology Bulletin*, *28*(4), 546-558. https://doi.org/10.1177/0146167202287011

Polivy, J., & Herman, C. P. (2002). If at first you don't succeed. False hopes of self-change. *American Psychologist*, *57*(9), 677-689.

Rauthmann, J. F., Horstmann, K. T., & Sherman, R. A. (2020). The psychological characteristics of situations: Towards an integrated taxonomy. In *The Oxford handbook of psychological situations.* (pp. 389-403). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.19

Rebele, R. W., Koval, P., & Smillie, L. D. (2021). Personality-informed intervention design: Examining how trait regulation can inform efforts to change behavior. *European Journal of Personality*, *35*(4), 623-645. https://doi.org/10.1177/08902070211016251

Reid, A., & Deaux, K. (1996). Relationship between social and personal identities: Segregation or integration. *Journal of Personality and Social Psychology*, *71*(6), 1084-1091. https://doi.org/10.1037/0022-3514.71.6.1084

Richard, F. D., BondJr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331-363. https://doi.org/10.1037/1089-2680.7.4.331

Ross, C. A., & Pam, A. (2005). *Pseudoscience in biological psychiatry: Blaming the body*. John Wiley & Sons.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 173-220). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-2601(08)60357-3

Rousis, G. J., Martel, F. A., Bosson, J. K., & Swann, W. B., Jr. (2023). Behind the blackpill: Self-verification and identity fusion predict endorsement of violence against women among self-identified incels. *Personality & Social Psychology Bulletin*, 1461672231166481. https://doi.org/10.1177/01461672231166481

Schein, E. H. (1956). The Chinese indoctrination program for prisoners of war. *Psychiatry*, *19*(2), 149-172. https://doi.org/10.1080/00332747.1956.11023044

Schiavone, S. R., & Vazire, S. (2022). Reckoning with our crisis: An agenda for the field of social and personality psychology. *Perspectives on Psychological Science*, *18*(3), 710-722. https://doi.org/10.1177/17456916221101060

Schlenker, B. R. (1980). *Impression Management : The Self-concept, Social Identity, and Interpersonal Relations*. Brooks/Cole.

Schlenker, B. R., & Leary, M. R. (1985). Social anxiety and communication about the self. *Journal of Language and Social Psychology*, *4*(3-4), 171-192. https://doi.org/10.1177/0261927x8543002

Sedikides, C., Skowronski, J. J., & Dunbar, R. I. M. (2006). When and why did the human self evolve? In *Evolution and social psychology.* (pp. 55-80). Psychosocial Press.

Shkurko, Y. (2019). Standard Social Science Model. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science* (pp. 1-9). Springer International Publishing. https://doi.org/10.1007/978-3-319-16999-6_1299-1

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487-510. https://doi.org/10.1146/annurev-psych-122216-011845

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). RETRACTED: Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, *109*(38), 15197-15200. https://doi.org/10.1073/pnas.1209746109

Siev, J., Zuckerman, S. E., & Siev, J. J. (2018). The relationship between immorality and cleansing: A meta-analysis of the Macbeth Effect. *Social Psychology*, *49*(5), 303-309. https://doi.org/10.1027/1864-9335/a000349

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366. https://doi.org/10.1177/0956797611417632

Simon, B. (2004). *Identity in Modern Society: A Social Psychological Perspective*. Blackwell Publishing.

Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science*, *30*(5), 711-727. https://doi.org/10.1177/0956797619831612

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power.

Journal of Personality and Social Psychology, *113*(1), 117-143. https://doi.org/10.1037/pspp0000096

Spears, R. (2001). The interaction between the individual and the collective self: Self-categorization in context. In *Individual self, relational self, collective self.* (pp. 171-198). Psychology Press.

Spiegel, A. (2016). Invisibilia In *The Personality Myth*.

Stake, J. E. (1992). Gender differences and similarities in self-concept within everyday life contexts. *Psychology of Women Quarterly*, *16*(3), 349-363. https://doi.org/10.1111/j.1471-6402.1992.tb00259.x

Stapel, D. A., & Lindenberg, S. (2011). RETRACTED: Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, *332*(6026), 251-253. https://doi.org/10.1126/science.1201068

Swann Jr, W. B. (1990). *To be adored or to be known? The interplay of self-enhancement and self-verification*The Guilford Press.

Swann, W. B. (2012). Self-verification theory. In *Handbook of theories of social psychology* (Vol. 2, pp. 23-42). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n27

Swann, W. B., & Jetten, J. (2017). Restoring agency to the human actor. *Perspectives on Psychological Science*, *12*(3), 382-399. https://doi.org/10.1177/1745691616679464

Swann, W. B., Jr., Gómez, Á., Seyle, D. C., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology*, *96*(5), 995-1011. https://doi.org/10.1037/a0013668

Swann, W. B., & Seyle, C. (2005). Personality psychology's comeback and its emerging symbiosis with social psychology. *Personality and Social Psychology Bulletin*, *31*(2), 155-165. https://doi.org/10.1177/0146167204271591

Tedeschi, J. T. (1981). *Impression Management Theory and Social Psychological Research*. Academic Press. https://doi.org/10.1016/B978-0-12-685180-9.50003-8

Thaler, R. H., & Sunstein, C. R. (2021). *Nudge : The final edition* (The final edition ed.). Penguin Books.

Turner, J. C. (1985). Social categorization and the self-concept: A social-cognitive theory of group behavior. In L. E. J. (Ed.), *Advances in group processes: Theory and research* (Vol. 2, pp. 77-122). JAI Press.

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, *20*(5), 454-463. https://doi.org/10.1177/0146167294205002

von Hippel, W., & Buss, D. M. (2017). Do ideologically driven scientific agendas impede the understanding and acceptance of evolutionary principles in social psychology? In L. Jussim & J. T. Crawford (Eds.), *Politics of Social Psychology* (Vol. 1). Psychology Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158-177. https://doi.org/10.1037/h0074428

Watson, J. B. (1925). *Behaviorism*. W.W. Norton & Company, Inc.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Front Psychol*, *7*, 1832. https://doi.org/10.3389/fpsyg.2016.01832

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*(5901), 606-607. https://doi.org/10.1126/science.1162548

Wilson, J., & Kelling, G. L. (1982). Broken windows: The police and neighbourhood safety. *The Atlantic Monthly*, *249*, 0-0.

Wood, J. V., Perunovic, W. Q., & Lee, J. W. (2009). Positive self-statements: power for some, peril for others. *Psychol Sci*, *20*(7), 860-866. https://doi.org/10.1111/j.1467-9280.2009.02370.x

Youyou, W., Yang, Y., & Uzzi, B. (2023). A discipline-wide investigation of the replicability of Psychology papers over the past two decades. *Proceedings of the National Academy of Sciences*, *120*(6), e2208863120. https://doi.org/10.1073/pnas.2208863120

Zhong, C.-B., Strejcek, B., & Sivanathan, N. (2010). A clean self can render harsh moral judgment. *Journal of Experimental Social Psychology*, *46*(5), 859-862. https://doi.org/https://doi.org/10.1016/j.jesp.2010.04.003

Zimbardo, P. G. (1969). *The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos*University of Nebraska Press.