AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

ISSN: 0022-3514

https://doi.org/10.1037/pspa0000450

Does Artificial Intelligence Cause Artificial Confidence? Generative Artificial Intelligence as an Emerging Social Referent

Taly Reich[1] and Jacob D. Teeny[2]
[1] Leonard N. Stern School of Business, New York University
[2] Kellogg School of Management, Northwestern University

As generative artificial intelligence (gen-AI) becomes more prevalent, it becomes increasingly important to understand how people psychologically respond to the content it explicitly creates. In this research, we demonstrate that exposure to gen-AI produced content can affect people's self-confidence at the same task through a social comparison process. Anchoring this research in the domain of creativity, we find that exposing people to creative content believed to have been created by gen-AI (vs. a human peer) *increases* people's self-confidence in their own relevant creative abilities. This effect emerges for jokes, stories, poetry, and visual art, and it can consequently increase people's willingness to attempt the activity—even though the greater confidence underscoring their actions might be unwarranted. We further show that these effects emerge because gen-AI is perceived as a lower social referent for creative endeavors, bolstering people's own self-perceptions. As a result, for domains in which gen-AI is perceived as an equal or greater social referent (i.e., in fact-based domains), the effects are attenuated. These findings have significant implications for understanding human–AI interactions, antecedents for creative self-confidence, and the known referents that people use for social comparison effects.

Statement of Limitations

The following studies demonstrate that exposure to creative content supposedly created by generative artificial intelligence (vs. a human) can enhance creative self-confidence via a social comparison process. Although these effects were tested and shown across different types and quality of creative content, several limitations should be noted. First, while the study designs prioritize internal validity—using very controlled stimuli and simplified paradigms—this may come at the expense of ecological validity. That is, the contexts in which participants encountered creative content were intentionally minimalistic; however, this may not have fully captured the complexity of real-world interactions with AI-generated material. Second, although our samples include participants from both the U.S. and the U.K., they are limited to Western cultural contexts. As such, the generalizability of these findings to non-Western populations remains an open question. Finally, our primary evidence for the proposed process relies on statistical mediation. While this approach robustly supported the theorized mechanism, valid criticisms of statistical mediation at large have been raised. We thus additionally address a key alternative explanation within our mediational analyses and provide additional support for the proposed process through a study using experimental moderation, which offers stronger causal evidence for the underlying psychological mechanism.

Keywords: artificial intelligence, self-confidence, creative output, social referent, generative artificial intelligence

Supplemental materials: https://doi.org/10.1037/pspa0000450.supp

Gandalf Nicolas served as action editor.

Taly Reich and Jacob D. Teeny contributed equally to this work. For all studies, they report all measures and conditions and have no data exclusions. Target sample sizes for all experiments were determined in advance of data collection based on the principle that researchers should collect at least 100 participants per condition (Simmons et al., 2018). Data collection did not continue after data analysis for any studies. The authors have made all preregistrations, study material, and data available on Researchbox (https://researchbox.org/3522&PEER_REVIEW_passcode=NRWMKY).

Taly Reich played a lead role in conceptualization and data curation and a supporting role in writing–review and editing. Jacob D. Teeny played a lead role in writing–review and editing.

Correspondence concerning this article should be addressed to Taly Reich, Leonard N. Stern School of Business, New York University, Tisch Hall, 40 West Fourth Street, New York, NY 10012, United States. Email: taly.reich@stern.nyu.edu

One of the most consequential efforts of the psychological sciences has been understanding how transformative technologies, such as internet search (e.g., Ward, 2021; Wegner & Ward, 2013) and smartphones (Horwood & Anglim, 2019; Miller, 2012), have altered people's self-conceptions and behaviors. In recent years, the public has been introduced to a flurry of new technologies with similar potential—non-fungible tokens, blockchain, the metaverse. However, none have had the staying power or prevalence following their introduction like the technology known as generative artificial intelligence (gen-AI).

Gen-AI, like ChatGPT, Gemini, Perplexity, and more, is a machine-based algorithm that can produce text, visuals, and audio that closely mimics the kind of content that humans produce (e.g., artwork, stories, information summaries; Lily et al., 2023). Although the familiarity with and capabilities of gen-AI will expand over time, understanding people's current responses provides a valuable foundation for future research. This is consistent with early, seminal research into the psychology surrounding internet search and smartphone usage (e.g., Kraut et al., 1998; Lee et al., 2014; McKenna & Bargh, 2014; Ward et al., 2017), where such theoretical groundwork offered useful insight even after the technologies had evolved. In the present work, we aim to build upon this tradition, asking how people's exposure to explicitly gen-AI created content affects their self-views and subsequent behavior. With gen-AI's growing presence in commercial messaging, internet culture, and beyond (e.g., Eloundou et al., 2024; Harreis et al., 2023; Matz et al., 2024), understanding how exposure to its content shapes people's perceptions—compared to identical content from humans—becomes increasingly important. Ultimately, we identify a robust phenomenon whereby exposure to gen-AI produced content can influence one of people's most influential views: their *self-confidence* in their own ability at the same task.

Bridging the literatures on social cognition and human–computer interactions, the present work shows how exposure to gen-AI's *creative* content (e.g., humor, stories, artwork) directly augments people's self-confidence at the same task. Across seven primary experiments (six preregistered) and four supplementary experiments ($N = 6{,}801$), when identical creative content is believed to stem from gen-AI (vs. a human peer), people have greater confidence in their own ability at that activity. Consequently, this can lead to a greater willingness to attempt that activity, even though their greater confidence might be unwarranted. We explain these effects through social comparison theory (Festinger, 1954). Whereas humans were previously the only social referent in this domain (Sun et al., 2024), gen-AI can serve as a relatively lower social referent, in turn bolstering people's self-confidence at the task. Based on this process, we show that this self-confidence effect is robust to differences in the content's quality (i.e., it emerges for a priori high- and low-quality work), and we identify an important boundary condition. The effect is restricted to domains where gen-AI serves as a lower social referent (creative domains) and is attenuated when it serves as an equal or greater one (fact-based domains).

Generative AI and Creative Self-Confidence

When new technology emerges, history finds that it often affects human society before those effects are fully realized, making research into the early psychological consequences of exposure to gen-AI content so important. In particular, we examine these consequences with respect to people's creative self-confidence (i.e., their self-perceived ability to perform creative tasks or activities relative to others; Karwowski & Barbot, 2016; Tierney & Farmer, 2002). Research on creative self-confidence has been of central interest to educators and organizational managers as it can affect a person's ability to adapt and learn (e.g., Ginns et al., 2023) as well as their motivation to innovate (e.g., Walumbwa et al., 2018). Indeed, this longstanding interest in studying the antecedents and moderators of creative self-confidence stems in part from the fact that this self-view plays a critical role in a person's willingness to even attempt creative tasks or goals in the first place (Tierney & Farmer, 2011). Without this self-confidence, people often hesitate to undertake relevant endeavors or might end them prematurely (Bandura, 1997).

In the present research, we examine how exposure to gen-AI produced creative content affects people's creative self-confidence, notably, by drawing on classic psychological theory surrounding social comparison processes.

Gen-AI as a Social Referent

Humans largely self-evaluate their proficiency at a task or skill as it stands relative to others (for a review, see Unkelbach et al., 2023). That is, people utilize others' perceived capability at relevant activities or actions to assess their own self-estimates of ability. The process by which these relative judgments are made—and the consequences they have for one's self-views—is comprised under *social comparison theory* (Festinger, 1954; Gerber et al., 2018).

Broadly speaking, there are three steps underlying social comparison processes (Mussweiler, 2003). First, a social referent (target or standard of comparison) is selected. Second, an attribute (both in the target and the perceiver) relevant to the judgment is identified. Third, accessible perceptions of the target's standing on the respective attribute are used to form a comparison point for subsequent self-evaluations. For example, a young woman might wonder how good she is at conversational storytelling. With her friends as accessible standards (Step 1), she might then consider how compelling their storytelling on average tends to be (Step 2), creating a comparison point for downstream self-evaluations of her own storytelling ability (Step 3). Earlier social comparison research largely focused on the first step (i.e., who perceivers selected as social referents; Festinger, 1954; Latané, 1966; Wheeler, 1991). However, more recent work has recognized that contexts often present a "forced social referent" (Gerber et al., 2018; Smith & Arnkelsson, 2000), where the target is already manifest (e.g., competing against someone or being exposed to someone's output). This then turns the research question to how people react to specific social referents, where we propose gen-AI is a novel (and now common) social referent.

Previous social psychological research has almost exclusively examined human referents in social comparison processes, because moderate perceptions of similarity (i.e., between the target and the perceiver) must exist for social comparison processes to initiate (see Collins, 2000). For the first time, gen-AI, which employs large-scale neural architectures (Sufi, 2024), allows for the nonhuman exhibition of humanlike capabilities in its production of coherent and contextually sensitive responses—what people perceive as "semi-human" (Baek et al., 2025; Bojić et al., 2024; De Vito, 2023; Lily et al., 2023). For example, people have already started turning to

gen-AI for relationship advice, conversing with it like another person (Marriott & Pitardi, 2024). These perceptions of gen-AI align with the conceptual framework introduced by Turing (1950) for when machines will be perceived like humans in terms of their ability to "think" (see also Searle, 1980). Moreover, given gen-AI's ability to personalize its content to people (Matz et al., 2024), this could further increase its perception as human (Schweitzer et al., 2024). Altogether, these first-of-its-kind technological capabilities of gen-AI make it a new candidate for social comparison judgments (i.e., a social referent) across a variety of self-comparisons.

Notably, and in line with the selective accessibility model of social comparison (Mussweiler, 2003), people would need to have gen-AI as a mentally accessible social referent to elicit such processes. As contextual exposure to a social referent is one of the surest ways to elicit mental accessibility (Mussweiler, 2003), we surveyed social media users (*Prolific*, $N = 150$, $M_{age} = 39.16$, $SD = 11.08$) on the extent to which the percentage of content they observe on social media was believed to be created by gen-AI (see the Supplemental Web Appendix for details). More specifically, we were interested in a domain of content for which humans have previously served as the sole form of social benchmark: creative output. To that effect, we inquired about the perceived percentage of gen-AI produced humorous content and visual art (two varied forms of creative content) that they observe. On average, participants indicated that they believed 29.33% of the humor and 42.40% of the visual art they observed on social media was produced by gen-AI. Thus, irrespective of participants' objective accuracy, gen-AI's perceived prevalence means that it can serve as a common social referent for one's creative abilities (Mussweiler, 2003; Smith & Arnkelsson, 2000). In the present research, we offer some of the first work examining how this new social referent for creativity affects people's creative self-confidence.

Perceptions of Creative Self-Confidence

Previous research has well established the importance of creative self-confidence for people's personal (Karwowski, 2016; Karwowski & Barbot, 2016; Lemons, 2010; Royston & Reiter-Palmon, 2019) and professional lives (Beghetto, 2006; Ginns et al., 2023; Puente-Díaz & Cavazos-Arroyo, 2017; Tierney & Farmer, 2011; Walumbwa et al., 2018). For example, higher creative self-confidence in students (Royston & Reiter-Palmon, 2019) and employees (Carmeli & Schaubroeck, 2007) was associated with greater engagement and involvement in creative tasks and problem solving. This body of research has therefore led to much interest in the antecedents to creative self-confidence (e.g., Puente-Díaz & Cavazos-Arroyo, 2017). We contribute to this literature by identifying a contextual antecedent: social comparison to gen-AI's creative content. How such exposure would affect creative self-confidence, though, relies on understanding whether gen-AI serves as a higher or lower social referent in this domain.

Returning to our earlier example, if a young woman recently heard her friends share stories (a forced social referent), she might then judge how compelling her own storytelling is relative to theirs. If she perceives her friends' stories to be more compelling (they serve as a higher social referent), this entails an *upward social comparison*. If she perceives them to be less compelling (a lower social referent), this entails a *downward social comparison*. Research has established that both types can subsequently lead to higher or lower self-judgments (Gerber et al., 2018; Suls et al., 2002; Wood, 1989); however, when the social referent is believed to be "moderately" similar to the perceiver, upward comparisons typically produce lower self-judgments (e.g., lower self-confidence at storytelling), and downward comparisons produce higher ones (e.g., higher self-confidence at storytelling; Collins, 2000; Morse & Gergen, 1970). Although gen-AI is distinct from humans in its lack of a corporeal form (Wang et al., 2015), it nonetheless shares a moderate degree of overlap in how people perceive it relative to humans (Santoro & Monin, 2023). Moreover, given its ability to seemingly "think," communicate, and produce content (Lily et al., 2023) alongside the humanizing ability to personalize its content (Schweitzer et al., 2024), gen-AI likely serves as a moderately similar social referent. This begs the question, then, of whether gen-AI is a higher or lower social referent for creative abilities in order to understand whether comparisons to it will lead to lower or higher creative self-confidence, respectively.

On the one hand, gen-AI could be seen as a higher social referent given its ability to produce coherent creative content at speeds faster and with greater diversity than any individual human. If so, gen-AI would generally lower perceivers' creative self-confidence. However, we argue that people carry fundamental perceptions about gen-AI, which suggest the opposite effects. Creativity is viewed as a core aspect of human identity (Haslam et al., 2013; Morewedge, 2022), deeply tied to imagination, emotion, and personal experience. Humans tend to see their creative abilities as uniquely human traits, one of the key factors that sets them apart from machines (Chatterjee, 2014). These perceptions could lead individuals to hold a general naïve theory that gen-AI, at a baseline, lacks the depth, intentionality, and/or personal touch to produce compelling creative work—even if it can produce creative content in mass quantity. Thus, people might be inclined to perceive gen-AI (relative to a human peer) as a lower social referent at creative tasks. This would mean that after viewing the exact same creative content—a joke, a story, a piece of art—attributing it to gen-AI (vs. a human) would augment self-confidence.

Notably, in assessing people's self-confidence, it can either be done with absolute or comparative measures. Absolute measures capture a person's self-perceived capability to successfully complete a task, whereas comparative measures capture a person's self-perceived ability relative to others (Freund & Kasten, 2012; Smith & Arnkelsson, 2000). By definition, social comparison effects are more aligned with the latter type of measure. Moreover, people themselves are more likely to make comparative (vs. absolute) self-judgments in contexts where objective ability is hard to determine (Festinger, 1954; Goffin & Olson, 2011). Thus, because evaluating the quality of creative work is largely considered a subjective practice (e.g., Cheek et al., 2021; Siev & Teeny, 2024; Stemler & Kaufman, 2020), comparative measures should be particularly ecologically relevant for assessing creative self-confidence. Altogether, then, we assess our outcome with comparative (vs. absolute) measures—which additionally have been argued to offer judgments that better correspond with people's self-perceptions and behavior (Freund & Kasten, 2012; Goffin & Olson, 2011). Nevertheless, we conduct supplementary studies (see the Supplemental Web Appendix) that show either type of measure produces the effect, and we discuss limitations of our primary comparative approach in the General Discussion section.

Consequences and Potential Boundary Conditions

As noted, comparative judgments of self-confidence can have a profound effect on people's relevant, everyday behaviors (see Goffin & Olson, 2011). With creative self-confidence specifically, there is correlational evidence that those with greater creative self-confidence are more likely to pursue creative endeavors (Tierney & Farmer, 2002). This suggests that if creative self-confidence is causally augmented (e.g., via exposure to gen-AI [vs. human] labeled work), it could increase people's likelihood to engage in relevant, creative activities. If true, it is worth noting that because this heightened self-confidence stems from a comparative perception, people may feel emboldened to attempt relevant tasks without their underlying abilities being changed (e.g., Burson, 2007). This could mean, on the one hand, that this effect could produce overconfidence and result in potentially negative outcomes (e.g., publicly putting forth creative content that is poorly received). On the other hand, this augmented confidence could be valuable for motivating attempts at creative activities or overcoming "blank page syndrome" (i.e., the difficulty to start a creative project due to low self-confidence; Balmer & Murcott, 2020). We will test whether this enhanced self-confidence can influence behavior and whether that behavior is objectively superior; however, we will only speculate on the normative value of such effects.

In addition to examining this fundamental consequence of creative self-confidence, we also test two potential boundary conditions, one we believe the effect to be robust to and another where we expect attenuation. First, we expect that this effect should emerge irrespective of the creative content's actual quality. Although creative work is inherently subjective, there nonetheless exists a continuum upon which people on average perceive it to be higher versus lower in quality (Stemler & Kaufman, 2020). However, because people use generalized perceptions of the target's relevant attribute to make judgments about their own standing (Mussweiler, 2003), we suspected exposure to gen-AI (vs. human) produced content would increase creative self-confidence regardless of the output's quality. That is, any individual might produce work that is above or below the quality of what they typically produce; therefore, to understand one's own standing at an ability, a perceiver needs to compare themselves to general perceptions of the target (e.g., Denrell & Liu, 2012; Reeder et al., 2001). This is not to say that the work's quality cannot (or would not) also affect a perceiver's self-confidence at the task. However, a social comparison process suggests that above and beyond any effect of the content's quality, generalized perceptions of gen-AI (vs. humans) as being less creative should result in greater reports of creative self-confidence after exposure to higher or lower quality output.

Second, and building on the point above, we do not expect gen-AI's effect on self-confidence to emerge across all domains. Generalized perceptions of a target's standing on a given attribute will depend on the domain to which that attribute applies (Zell & Krizan, 2014). In the case of gen-AI, although it is generally perceived to be lower in standing for attributes related to creative ability, it is generally perceived to be higher in standing for attributes related to fact-based abilities. That is, people generally evaluate gen-AI's competency at tasks that rely on factual accuracy quite positively (Castelo et al., 2019). For example, gen-AI is expected to generate summaries of information or provide knowledge-based answers with a high degree of accuracy (Korteling et al., 2021). Thus, relative to creative domains where gen-AI serves as a lower social referent, with respect to fact-based domains, gen-AI should serve as an equal or higher social referent. In the case of a moderately similar target, then, the confidence-enhancing effect of exposure to gen-AI (vs. human) content should attenuate or even disappear for fact-based domains.

Finally, and as described at the outset, the effects documented in this research could be affected by people's familiarity with gen-AI—either as it stands now or as it develops into the future. Although we explore how differences in current experiences with gen-AI affect our primary phenomenon, we provide some speculatory data and commentary in the General Discussion section about whether or how these effects might evolve in the future.

Research Overview

For the following studies, Studies 1B, 1C, 2, 3, 4, and 5 were preregistered. All data, preregistrations, and materials are posted here (https://researchbox.org/3522). In all studies, we recruit 200 participants per condition, which provide .80 power (two-tailed, $\alpha = .05$) to detect an effect as small as $d = .28$ in two-cell designs and $d = .20$ for main effects in 2×2 designs. All participants were collected in a nonoverlapping manner between June 2024 and February 2025, were paid $0.20 per minute of their time (average duration: 1–3 min per study), and no participants were excluded from analyses.[1] With the exception of Study 3 and one supplemental study, in which participants were recruited from the United Kingdom, participants were recruited from the United States with English as their first language and desktop computers as the devices on which they could take the survey. Additional analyses and commentary (e.g., the zero-order correlation between all primary measures) can be found in the Supplemental Web Appendix.

Studies 1A–1C show that after exposure to the same content labeled as being produced by gen-AI (vs. a human peer), participants report greater creative self-confidence in their relevant abilities (i.e., humor, 1A; visual art 1B; poetry, 1C). All studies additionally provide mediational evidence for the proposed social comparison process, whereby participants perceive gen-AI (vs. a human peer) to have lower ability in that creative domain (i.e., serve as a lower social referent). Study 2 uses a new domain (storytelling) and demonstrates how this confidence effect can translate into a greater willingness to engage in the task, while Study 3 shows how enhanced confidence does not lead to enhanced objective ability. Studies 4 and 5 test potential theory-based boundary conditions. In Study 4, participants demonstrate enhanced creative confidence irrespective of the content's actual quality, and in Study 5, the effect is attenuated when the domain is changed to one where gen-AI serves as an equivalent social referent (i.e., fact-based summarizing). Finally, we provide additional data in the Supplemental Web Appendix, which is briefly described where applicable in our empirical section as well as the General Discussion section.

[1] Although we retain all participants in our analyses to maximize power and prevent ad hoc exclusions, additionally constraining the sample to more attentive/engaged participants (i.e., removing duplicate IP addresses; removing those with impossibly short survey completion times) retains all effects and their significance.

Studies 1A, 1B, 1C: Creative Self-Confidence Across Domains

Studies 1A–1C test how labeling the exact same creative content as being generated by gen-AI (vs. a human peer) affects participants' self-perceptions of their own ability to engage in that activity. Each study uses a different type of creative output and presents the content in a minamalist paradigm (e.g., analgous to how content is shared in a direct message, email, or other form of digital communication) in order to test the robustness of the effect while ensuring the highest internal validity. Each study additionally tests our proposed social referent process, whereby people perceive gen-AI (vs. humans) as worse producers of creative content, which leads participants to have greater confidence in their own abilities at the task.

Study 1A: Method

Participants and Procedure

We recruited 400 participants from *Prolific* (M_{age} = 36.65, *SD* = 12.39, male = 37.8%, female = 60.8%, other = 1.5%) who were randomly assigned to a two-cell (source: gen-AI vs. human) between-subjects design.

Participants began by learning that either a Prolific survey taker or a generative AI was asked to write a joke about the antidepressant Lexapro. In reality, we used the gen-AI Perplexity (2024) to write the joke: "Why was the Lexapro pill feeling down? Because it had a depressing job! But don't worry, it took itself and now it's doing much better." Next, participants reported their *creative self-confidence* ("To what extent do you think you could have come up with a better joke?" 1 = *not at all*, 7 = *extremely*) and *perceptions of the source's ability* ("To what extent does the joke writer have a sense of humor?" 1 = *not at all*, 7 = *extremely*).

Study 1A: Results

A 2-cell analysis of variance (ANOVA; source: gen-AI vs. human peer) revealed a significant effect on participants' creative self-confidence. When the same joke was labeled as coming from a gen-AI (vs. human peer), participants reported greater confidence in their abilities to devise a better joke, $M_{gen\text{-}AI}$ = 3.71, *SD* = 1.79 versus M_{human} = 3.24, *SD* = 1.74; $F(1, 398) = 7.08$, $p = .008$, 95% CI [0.123, 0.818]; $d = 0.27$. We also observed that gen-AI was perceived to have lower abilities at generating humorous content, $M_{gen\text{-}AI} = 3.94$, $SD = 1.67$ versus $M_{human} = 4.57$, $SD = 1.51$; $F(1, 398) = 15.81$, $p < .001$, 95% CI [−0.945, −0.320]; $d = -0.40$. A test of statistical mediation (Hayes, 2022; Model 4) supports the idea that perceiving gen-AI as less able to generate humor is associated with participants' greater creative self-confidence ($b = .12$, $SE = .05$, 95% CI [0.037, 0.234]).[2]

Study 1B: Method

Participants and Procedure

This study was preregistered (https://aspredicted.org/5YV_4TM). We recruited 400 participants from *Prolific* and received 399 (M_{age} = 35.77, *SD* = 12.15, male = 39.1%, female = 58.9%, other = 2.0%) randomly assigned to a two-cell (source: gen-AI vs. human) between-subjects design.

This study examined participants' self-confidence in their ability to draw visual art. Participants began by learning that either a Prolific survey taker or a gen-AI had been asked to draw an "image symbolizing protection." Participants were then shown a shaded and detailed pencil-drawing of an open hand with an eye in its palm. Participants were then asked to indicate "how confident" and "how certain" they "could have come up with a better drawing" (1 = *not at all*, 7 = *extremely*). These two items were averaged ($r = .89$, $p < .001$). Participants were additionally asked the extent to which the "creator of this piece" is a good drawer (1 = *not at all*, 7 = *extremely*).

Study 1B: Results

An ANOVA (source: gen-AI vs. human peer) once more revealed the same pattern as Study 1A: exposure to art labeled as coming from a gen-AI (vs. human peer) resulted in greater self-confidence in participants' drawing abilities, $M_{gen\text{-}AI} = 4.07$, $SD = 2.02$ versus $M_{human} = 3.00$, $SD = 1.79$; $F(1, 397) = 31.47$, $p < .001$, 95% CI [0.695, 1.45]; $d = 0.56$. We also observed the same pattern as Study 1A on perceptions of the author's ability, $M_{gen\text{-}AI} = 4.63$, $SD = 1.64$ versus $M_{human} = 5.78$, $SD = 1.12$; $F(1, 397) = 66.83$, $p < .001$, 95% CI [−1.42, −0.871]; $d = -0.82$, and this difference exhibited a pattern consistent with statistical mediation (Hayes, 2022; Model 4) for the effect on participants' self-confidence ($b = .44$, $SE = .10$, 95% CI [0.263, 0.639]).

Study 1C: Method

Participants and Procedure

This study was preregistered (https://aspredicted.org/tryz-4jdh.pdf). We recruited 400 participants from *Prolific* (M_{age} = 39.87, *SD* = 13.42, male = 49.0%, female = 49.0%, nonbinary = 1.3%, prefer not to say = 0.8%) who were randomly assigned to a two-cell (source: gen-AI vs. human) between-subjects design.

This study examined participants' self-confidence in their ability to write a poem. As in our prior studies, participants learned that either a Prolific survey taker or a gen-AI was the author of the poem. Participants then viewed a lesser known but highly rated poem ("Snow Drift") by esteemed American poet Robert Frost. After reading the poem, participants reported their creative self-confidence on the same two items as Study 1B, albeit tailored to their ability at poetry ($r = .90$, $p < .001$), as well as their perceptions of the author's writing abilities. In this study, participants additionally reported their familiarity with gen-AI's capabilities (1 = *not at all familiar*, 7 = *very familiar*) and their exposure to gen-AI content (1 = *not at all*, 7 = *a great deal*). We use these measures to conduct additional analyses to determine

[2] Across our studies, our primary evidence for process comes through statistical mediation, which has been criticized for its overreliance on model fit, potential for measurement error, and susceptibility for overstated claims (Fiedler et al., 2018). In the present research, we simply present the mediational relationship as consistent support for the proposed associations theorized in our data. We later bolster this evidence with a process-by-moderation experimental design (Spencer et al., 2005), while nonetheless leaving open that other psychological processes could be involved in the effect (see the General Discussion section for such considerations).

how preestablished experiences with gen-AI might influence the results.

Study 1C: Results

An ANOVA (source: gen-AI vs. human peer) replicated the pattern of effects so far observed. Exposure to the same poem ostensibly authored by gen-AI (vs. human peer) resulted in greater creative self-confidence at writing poetry, $M_{\text{gen-AI}} = 3.74$, $SD = 1.74$ versus $M_{\text{human}} = 2.91$, $SD = 1.65$; $F(1, 398) = 23.99$, $p < .001$, 95% CI [0.498, 1.17]; $d = 0.49$. We also observed the same pattern as Studies 1A and 1B on perceptions of the author's ability, $M_{\text{gen-AI}} = 3.88$, $SD = 1.52$ versus $M_{\text{human}} = 4.56$, $SD = 1.39$; $F(1, 397) = 21.49$, $p < .001$, 95% CI [−0.964, −0.390]; $d = -0.47$, and this difference statistically mediated (Hayes, 2022; Model 4) the effect of gen-AI's label on participants' greater self-confidence ($b = .22$, $SE = .07$, 95% CI [0.104, 0.358]).

As an additional set of analyses, we also reran the aforementioned ANOVAs as analyses of covariance, controlling for participants' a priori familiarity with and exposure to gen-AI content. In doing so, all effects and their significance held (ps < .001). Among those exposed to gen-AI content, we also examined the relationship between participants' familiarity with and/or exposure to gen-AI and their creative self-confidence. For both control measures, the relationships were nonsignificant (familiarity, $b = -.04$, $p = .615$ exposure, $b = -.11$, $p = .157$).

Study 2: Gen-AI Effect on Willingness to Engage

Study 2 extends our findings by examining a new creative domain (storytelling) as well as a consequence of peoples' greater creative self-confidence from observing gen-AI (vs. human) creative content: their willingness to engage in the relevant activity. Online forums (e.g., Reddit's r/WritingWithAI) host user-generated short stories from both humans and gen-AI, and this study uses a carefully controlled paradigm to test how such exposure might affect readers' self-confidence at and willingness to engage in creative writing. Moreover, by collecting a measure of writing willingness, we strengthen the construct validity of our self-confidence items and test whether effects on it can translate to behavioral outcomes.

Method

Participants and Procedure

This study was preregistered (https://aspredicted.org/J9Y_426). We recruited 400 participants from *Prolific* ($M_{\text{age}} = 37.48$, $SD = 12.22$, male = 45.3%, female = 52.8%, other = 2.0%) randomly assigned to a two-cell (source: gen-AI vs. human) between-subjects design.

Participants were initially told that either a Prolific survey taker or a gen-AI was asked to write a story using a story spine (i.e., a sequentially prompted outline). This story spine first asked the author to provide a sentence starting with "Once upon a time," followed by another sentence starting with "Every day," and then another with "Until one day," and so forth, until a complete story was outlined. In the present study, the author (gen-AI or a human peer) generated a story about rain, which described how rain brought both excitement and sickness to a town. Participants then responded to the same self-confidence items ($r = .93$, $p < .001$) and perceptions of the author's storytelling abilities as Study 1C. Novel to this study, participants were led to believe that after the study, they would have the opportunity to write a story using the story spine, and we assessed their behavioral willingness to engage in this task (1 = yes, 2 = no). Given that Prolific workers are motivated to finish the survey as reasonably quickly as possible (Ipeirotis, 2010), this measure serves as a conservative test of our theory. At the same time, because participants' time is valuable to them, those who elected to use the story spine were not required to do so and could move on to the end of the survey.

Results

An ANOVA (source: gen-AI vs. human peer) revealed a significant effect on participants' creative self-confidence. When the same story was labeled as coming from a gen-AI (vs. human peer), participants reported greater self-confidence in their ability to write a better story, $M_{\text{gen-AI}} = 5.00$, $SD = 1.64$ versus $M_{\text{human}} = 3.90$, $SD = 1.66$; $F(1, 398) = 44.99$, $p < .001$, 95% CI [0.783, 1.43]; $d = 0.67$. Again, we observed that the gen-AI (vs. human peer) author was perceived to have lower abilities at writing stories, $M_{\text{gen-AI}} = 2.86$, $SD = 1.42$ versus $M_{\text{human}} = 3.77$, $SD = 1.40$; $F(1, 398) = 41.58$, $p < .001$, 95% CI [−1.19, −0.633]; $d = -0.65$, which, in line with Studies 1A–1C, mediated the effect on self-confidence ($b = .12$, $SE = .05$, 95% CI [0.783, 1.43]).

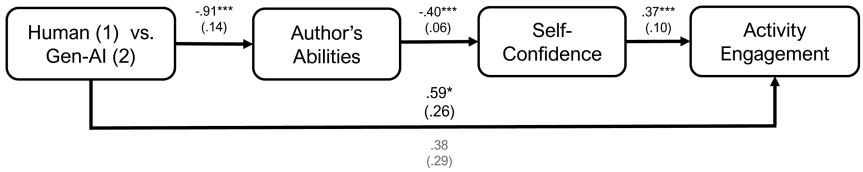
Next, we looked at participants' willingness to use the story spine themselves. A chi-square analysis revealed a significant effect ($\chi^2 = 5.26$, $p = .022$, Cramer's $V = .12$). Whereas only 14.5% of participants were willing to write a story after seeing a human write one, 23.5% were willing to after it was ostensibly written by gen-AI. We next conducted a test of sequential mediation (Hayes, 2022; Model 6), where we examined whether exposure to gen-AI (vs. human) labeled content increased willingness to use the story spine because (a) people perceived gen-AI (vs. a human peer) to have worse abilities at storytelling, which (b) led to greater self-confidence in their own abilities at storytelling, increasing their willingness to engage in the task (see Figure 1). This indirect effect was significant ($b = .13$, $SE = .05$, 95% CI [0.060, 0.248]). Notably, reversing the serial mediators in this pattern of analyses renders the indirect effect nonsignificant.

Study 3: Objective Differences in People's Creative Output

In replicating our focal effects on new creative content (a captioned comic), Study 3 also has all participants generate creative content themselves. That is, similar to how people post humorous content via online memes or even participate in comic generation contests (e.g., like the cartoon captioning contest at *The New Yorker*), all participants were subsequently asked to caption the comic themselves. We then had external coders rate the quality (i.e., funniness) of the caption they generated to test whether the self-confidence elicited by the source (gen-AI vs. human) of the preceding caption was merited. In other words, did the greater self-confidence of those exposed to gen-AI content actually translate into greater ability at the relevant activity?

Figure 1

Serial Mediation in Study 2



Note. The figure represents the serial mediation of source (human vs. gen-AI) on participants' perceptions of the ostensible author's ability in the domain, participants' self-confidence in their own abilities at it, and then their willingness to engage in a relevant activity. In the figure, the top numbers represent the unstandardized betas, and the numbers in parentheses represent their standard error. The numbers in gray represent the descriptive statistics associated with the direct effect. gen-AI = generative artificial intelligence.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Method

Participants and Procedure

This study was preregistered (https://aspredicted.org/TTV_LK2). We recruited 400 participants from *Prolific*, receiving 401 (M_{age} = 40.65, SD = 14.604, male = 57.9%, female = 40.9%, nonbinary = 1.2%) who were randomly assigned to a two-cell (source: gen-AI vs. human) between-subjects design.

Similar to Study 1A, participants began by seeing a cartoon drawing for which either a Prolific survey worker or gen-AI had supposedly been asked to write a funny caption (i.e., a mouse pointing a revolver at a cat, saying: "You want a piece of me, pussycat? I'll go full Clint Eastwood on your furry behind."). Participants then reported their self-confidence ($r = .92$, $p < .001$) as well as their perceptions of the author's sense of humor on the same items as Study 2. Afterward, participants were asked to generate their own caption for the cartoon. Once they had, we asked them to rate how funny it was on two items ("is funny [made you laugh]" 1 = *not at all*, 7 = *extremely*; $r = .87$, $p < .001$). Notably, this index serves to provide further validity for our primary self-confidence items in more of an absolute (vs. comparative) form (for more evidence of the effect on other forms of absolute measures, see the Supplemental Web Appendix).

After data collection, we recruited three research assistants (intercoder reliability: $\alpha = .632$) who were blind to the hypothesis and condition to rate each of these captions on the same measures (i.e., how funny and made them laugh, 1 = *not at all*, 7 = *extremely*).

Results

First, we note that an ANOVA (source: gen-AI vs. human) replicated our primary findings. Those exposed to gen-AI (vs. human) content reported greater self-confidence in their own ability at the task, $M_{gen\text{-}AI} = 3.77$, $SD = 1.62$ versus $M_{human} = 3.05$, $SD = 1.59$; $F(1, 399) = 19.73$, $p < .001$, 95% CI [0.397, 1.03]; $d = 0.45$. We also observed the consistent pattern on perceptions of the author's ability, $M_{gen\text{-}AI} = 3.85$, $SD = 1.78$ versus $M_{human} = 4.56$, $SD = 1.46$; $F(1, 399) = 19.49$, $p < .001$, 95% CI [−1.04, −0.398]; $d = -0.44$, and this difference statistically mediated the effect on self-confidence ($b = .09$, $SE = .05$, 95% CI [0.010, 0.192]).

Next, we tested how funny participants found their own captions to be. Mirroring the results on our comparative self-confidence index, those who were first exposed to gen-AI (vs. human) content rated their own captions as funnier, $M_{gen\text{-}AI} = 3.52$, $SD = 1.70$ versus $M_{human} = 3.16$, $SD = 1.55$; $F(1, 399) = 4.71$, $p = .031$, 95% CI [0.033, 0.673]; $d = 0.22$. However, when external coders were asked to rate the funniness of the captions, they found no differences ($M_{gen\text{-}AI} = 1.92$, $SD = 0.74$ vs. $M_{human} = 1.98$, $SD = 0.74$; $F < 1$). In short, the greater confidence in one's abilities from exposure to gen-AI (vs. human) creative content may be unmerited.

Study 4: Manipulating Creative Content Quality

In our next study, we test an important factor in the prior effects: the quality of the output. Based on the selective accessibility model of social comparison, we should expect generalized perceptions of the target's relevant attribute to influence self-evaluations, above and beyond the content's quality. Although the content's quality can still impact self-evaluations (e.g., observing better content could result in lower self-confidence and vice versa), generalized perceptions that gen-AI (vs. humans) are less capable at creating creative work should continue to elicit higher creative self-confidence following exposure to its content.

Method

Participants and Procedure

This study was preregistered (https://aspredicted.org/CDG_17T). We recruited 800 participants from *Prolific*, receiving 802 (M_{age} = 37.72, SD = 13.11, male = 39.8%, female = 58.9%, nonbinary = 1.4%) who were randomly assigned to a 2 (source: gen-AI vs. human) × 2 (quality: lower vs. higher) between-subjects design.

This study consisted of two waves. First, we conducted a pretest ($N = 100$), where participants were shown four different captions generated by participants in Study 3 (i.e., for the cartoon of a mouse holding a revolver on a cat). Two captions selected for this pretest were a priori low in quality, while two others were high in quality. Participants individually rated all

of the captions in how funny they were (1 = *not at all*, 7 = *extremely*), where paired sample *t* tests identified the least and most funny captions, respectively: "Back up, cat" ($M = 2.58$, $SD = 1.64$) and "I have 9 lives. You have 6 bullets. Do the math, mouse," $M = 4.75$, $SD = 1.49$; $t(99) = 9.79$, $p < .001$, $d = 2.22$. In the focal study (Study 4), participants were shown the same illustration with either the low- or high-quality caption, and it was ascribed to either gen-AI or a Prolific survey taker. Afterward, participants reported their confidence in their abilities at creating a funnier caption ($r = .92$, $p < .001$) and their perception of the source's ability to generate humor more broadly on the same items as Study 3.

Results

We first conducted a 2 (source: gen-AI vs. human) × 2 (quality: low vs. high) ANOVA on participants' creative self-confidence. From this, we observed a main effect of source, where exposure to gen-AI (vs. human) labeled content led to greater self-confidence, $M_{\text{gen-AI}} = 3.61$, $SD = 1.80$ versus $M_{\text{human}} = 3.19$, $SD = 1.76$; $F(1, 798) = 13.44$, $p < .001$, 95% CI [0.197, 0.651]; $d = 0.24$. We also observed a main effect of quality, where participants reported greater creative self-confidence following the low- (vs. high-) quality content, $M_{\text{low}} = 4.11$, $SD = 1.73$ versus $M_{\text{high}} = 2.71$, $SD = 1.56$; $F(1, 798) = 146.17$, $p < .001$, 95% CI [−1.62, −1.17]; $d = 0.85$. These effects were not qualified by an interaction ($F < 1$), suggesting that people report greater creative self-confidence following exposure to gen-AI (vs. humans), irrespective of variations in the content's quality.

We next conducted the same 2 × 2 ANOVA on participants' perceptions of the author's ability to generate humor, where we observed a main effect of the source: A gen-AI (vs. human) author was perceived to have lower abilities at humor, $M_{\text{gen-AI}} = 3.85$, $SD = 1.85$ versus $M_{\text{human}} = 4.64$, $SD = 1.82$; $F(1, 798) = 48.80$, $p < .001$, 95% CI [−1.01, −0.569]; $d = -0.43$. We also observed a main effect of quality, where participants perceived the author to have a lower ability to generate humor following the low- (vs. high-) quality content, $M_{\text{low}} = 3.35$, $SD = 1.74$ versus $M_{\text{high}} = 5.13$, $SD = 1.55$; $F(1, 798) = 249.25$, $p < .001$, 95% CI [1.57, 2.01]; $d = -1.08$. These effects were not qualified by an interaction, $F(1, 798) = 1.47$, $p = .226$, again suggesting that gen-AI (vs. humans) is perceived to be less capable of generating humor, irrespective of the content's quality.

Finally, a test of mediation (Hayes, 2022; Model 4) supported our earlier findings: collapsing across high- and low-quality content, perceiving gen-AI (vs. humans) as less capable of generating humor was associated with greater self-confidence ($b = .24$, $SE = .05$, 95% CI [0.154, 0.347]).

Study 5: Moderation by Domain

So far, we have shown that people report greater creative self-confidence after observing ostensibly gen-AI (vs. human) creative content, because people perceive gen-AI to be a lower social referent (i.e., a downward comparison against a moderately similar target, which magnifies their own self-evaluations). Our final study varies not only the source (gen-AI vs. human) but also the type of content: fact-based versus creative-based. We should expect to replicate the effect observed previously for creative-based content (i.e., when gen-AI is a lower social referent); however, for fact-based content, we should expect no differences (i.e., when gen-AI is an equal social referent).

Method

Participants and Procedure

This study was preregistered (https://aspredicted.org/151_69W). We recruited 800 participants from *Prolific*, receiving 802 ($M_{\text{age}} = 42.12$, $SD = 13.93$, male = 42.1%, female = 57.2%, nonbinary = 0.6%) who were randomly assigned to a 2 (source: gen-AI vs. human) × 2 (domain: creative-based vs. fact-based) between-subjects design.

All participants read content described as coming from either gen-AI or a Prolific survey worker on the topic of rain. For half of these participants, their content came in the form of a creative domain, a story spine about rain (the same as Study 2). For the other half of participants, their content came in the form of factual information about rain. This paragraph (which was approximately the same in length as the story spine) briefly described what rain was, how it formed, and some of its consequences (e.g., the distinct earthy smell it can release). Afterward, participants were asked to report their self-confidence ($r = .94$, $p < .001$) as well as their perceptions of the author's abilities on the same items from Study 2.

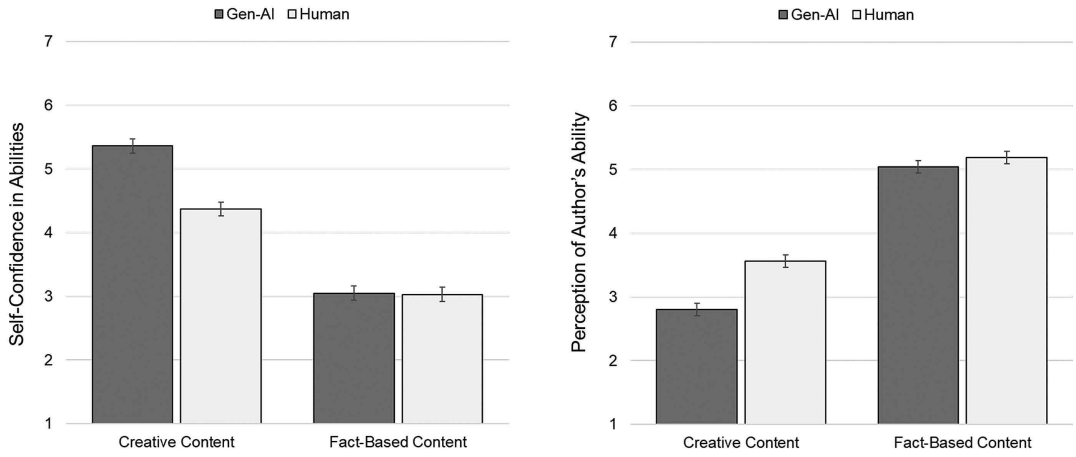
Results

We first conducted a 2 (source: gen-AI vs. human) × 2 (content: fact-based vs. creative-based) ANOVA on participants' creative self-confidence. From this, we observed a main effect of source, where exposure to gen-AI (vs. human) content again led to greater self-confidence in their own ability at the task, $M_{\text{gen-AI}} = 4.20$, $SD = 1.92$ versus $M_{\text{human}} = 3.70$, $SD = 1.79$; $F(1, 798) = 19.95$, $p < .001$, 95% CI [0.283, 0.726]; $d = 0.27$. We also observed a main effect of content, where participants reported greater self-confidence at producing better creative (vs. fact-based) content, $M_{\text{creative}} = 4.86$, $SD = 1.67$ versus $M_{\text{fact}} = 3.04$, $SD = 1.60$; $F(1, 798) = 260.49$, $p < .001$, 95% CI [1.60, 2.05]; $d = 1.11$. Most importantly, these effects were qualified by the predicted interaction, $F(1, 798) = 18.23$, $p < .001$; see Figure 2, left panel. Decomposing this by the type of content, we replicated our earlier studies on creative-based self-confidence, $M_{\text{gen-AI}} = 5.36$, $SD = 1.49$ versus $M_{\text{human}} = 4.37$, $SD = 1.70$; $F(1, 798) = 38.17$, $p < .001$, 95% CI [0.673, 1.30]; $d = 0.62$. However, we observed no difference in people's fact-based self-confidence as a function of the content's author ($M_{\text{gen-AI}} = 3.05$, $SD = 1.58$ vs. $M_{\text{human}} = 3.03$, $SD = 1.62$; $F < 1$).

We next conducted the same 2 × 2 ANOVA on perceptions of the author's ability to write their respective content. First, we note that gen-AI (vs. human) was overall perceived to be a worse writer, $M_{\text{gen-AI}} = 3.92$, $SD = 1.78$ versus $M_{\text{human}} = 4.37$, $SD = 1.53$; $F(1, 798) = 23.25$, $p < .001$, 95% CI [−0.642, −0.270]; $d = -0.27$. Additionally, the author was perceived as a worse writer when the content was creative- (vs. fact-) based, $M_{\text{creative}} = 3.18$, $SD = 1.46$ versus $M_{\text{fact}} = 5.11$, $SD = 1.27$; $F(1, 798) = 417.86$, $p < .001$, 95% CI [−2.12, −1.75]; $d = -1.41$. Most importantly,

Figure 2

Moderation by Domain in Study 5



Note. The graphs depict the effect of source (gen-AI vs. human) and the content domain (creative vs. fact-based) on participants' self-confidence (left panel) and perceptions of the author's ability in that domain (right panel). gen-AI = generative artificial intelligence.

these effects were qualified by the predicted interaction, $F(1, 798) = 10.81$, $p = .001$; see Figure 2, right panel. Whereas we replicate our earlier studies on perceptions of the author's ability for creative-based content, $M_{\text{gen-AI}} = 2.80$, $SD = 1.44$ versus $M_{\text{human}} = 3.56$, $SD = 1.37$; $F(1, 798) = 32.88$, $p < .001$, 95% CI [−1.03, −0.505]; $d = -0.54$, gen-AI and humans were perceived as equally capable in writing content when the domain was fact-based, $M_{\text{gen-AI}} = 5.04$, $SD = 1.31$ versus $M_{\text{human}} = 5.19$, $SD = 1.22$; $F(1, 798) = 1.18$, $p = .278$, 95% CI [−0.408, 0.117]; $d = -0.12$.

Finally, we tested a pattern of moderated mediation (Hayes, 2022; Model 7), where the overall significant effect of the source led to greater self-confidence through perceptions of the author's ability when a gen-AI (vs. human) wrote about creative (rather than fact-based) content. This index of moderated mediation was significant ($b = .40$ $SE = .12$, 95% CI [0.160, 0.647]), resulting in a significant indirect effect for creative content ($b = .49$, $SE = .09$, 95% CI [0.310, 0.677]) but not for fact-based content ($b = .09$, $SE = .08$, 95% CI [−0.607, 0.258]).

General Discussion

Understanding how people respond to transformative technologies like gen-AI helps us understand important but potentially unrecognized effects it can have on people's everyday psychology. In the present work, we found that observing creative content ostensibly generated by a gen-AI (vs. human) enhanced people's relevant, creative self-confidence, which could subsequently affect their behavior. We further provided evidence for the effect's process via a social comparison mechanism. People tend to perceive gen-AI as a lower social referent, and this downward comparison increases their self-confidence. Importantly, and in support of this process, these effects emerged above and beyond the creative content's quality and were diminished when gen-AI served as an equal social referent (in a fact-based domain).

Although we believe these findings offer valuable theoretical contributions for the literatures on social comparison (e.g., a novel, emerging social referent) and creative self-confidence (e.g., a novel antecedent), we also believe it carries important practical contributions. From survey data (McKinsey & Company, 2024) as well as our own pilot data, people are being increasingly exposed to gen-AI creative content. Whether this is a peer's gen-AI produced digital art or the stories from entirely gen-AI mega-influencers (e.g., @lilmiquela), the prevalence of this new social referent is robust. On the one hand, this might raise concerns about the potential for overconfidence in people's creative abilities; however, on the other, these effects offer insight for new strategies to help motivate engagement in creative tasks or pursuits. For example, in helping motivate students to write a personal essay for an application, exposure to gen-AI-created essays might do better to bolster their self-confidence at the task relative to seeing another student's prior work. The same could be true in occupational settings, such as finding the self-confidence to write a cover letter to a dream job or the self-confidence to design and give a presentation. In short, we believe the basic research conducted here could offer insight not only into the widespread consequences of people's increasing exposure to gen-AI creative content but also into strategies for leveraging this technology to further promote human creativity.

Gen-AI: A Unique Social Referent?

Prior to this work, most of the research on gen-AI has examined its capabilities (e.g., Eloundou et al., 2024), rather than how exposure to what it produces affects people's self-views and behavior. We have suggested that gen-AI's humanlike traits lead people to treat it as a social referent, making it the subject of social comparisons once reserved for human targets. This implies that gen-AI could serve as

a social referent for other phenomena (e.g., social comparisons on opinions and beliefs), raising interesting questions there, too. However, one area we believe to be particularly relevant to the current work is the fungibility of perceptions about gen-AI's creativity. From Study 1C, it appears that the naïve theory regarding gen-AI's creative ability is relatively stable, because neither prior familiarity with nor exposure to gen-AI altered the effect. However, this does not mean perceptions of gen-AI might not change over time. A social referent's ability can be conceived in terms of their current capabilities and/or their aptitude to improve (Smith & Arnkelsson, 2000). With gen-AI, though, it could be that creativity's fundamental connection to humanity means people perceive gen-AI as incapable of improving at creativity—a potentially unique property for this social referent. Put differently, might people be resistant to updating their beliefs about the general creative ability of gen-AI relative to a fellow human, thereby constraining the extent of its impact on self-evaluations?

To test this, we ran two additional studies (see the Supplemental Web Appendix for details). In both, we manipulated the source of the creative content (gen-AI vs. human) as well as a trait integral to producing creative work (Haslam et al., 2013): emotional depth ($N = 800$) or authenticity ($N = 800$). If creativity is something people believe to be fundamental only to humans, describing gen-AI as either more emotionally deep or authentic should not affect people's relevant self-confidence (e.g., gen-AI should remain a lower social referent to the same degree regardless of how emotionally deep it supposedly is). However, across both studies, this was not the case. Being exposed to the version of gen-AI higher (vs. lower) in emotional depth or authenticity (i.e., the version that acted as a higher social referent) resulted in lower creative self-confidence ($ps \leq .005$). Notably, this effect of the creativity manipulation was equally true for humans (i.e., there was no interaction between the source [human vs. gen-AI] and the creative trait [high vs. low]). This suggests that people can update their naïve theories about gen-AI'screativity similar to how they might update their perceptions of other human referents. In these studies, we also found that exposure to gen-AI (vs. human) content continued to result in greater creative self-confidence overall ($ps \leq .013$), pointing to the strength of people's general perceptions of gen-AI as a lower social referent to humans in this domain.

So, are there areas of social comparison processes that might be unique to gen-AI? Or is it another—albeit highly prevalent—social referent that will now operate similarly to other human targets? We believe one interesting area for future research is how gen-AI might be distinct from humans with respect to variations in similarity. The perceived similarity between the target and the individual plays a critical role in social comparison effects (see Collins, 2000). However, much of the prior work showcasing this has increased similarity via physical traits, identities, or geographies. Thus, many of these prior inductions might be inapplicable to gen-AI, limiting its ability to produce the range of social comparison effects observed for human targets. For example, in contexts of very high perceiver–target similarity, upward social comparisons can increase self-confidence (Mussweiler, 2003). However, given gen-AI's perception of being only "semihuman," such assimilative effects might be restricted to human referents.

Of course, with time, gen-AI might develop into something closer to humans themselves (e.g., via improved conversational dynamics, corporeal forms, etc.). In these instances, the effects and speculation presented in this research might change. Although some of these developments are still far into the future, one aspect relevant to this discussion is potentially much more proximal: greater quality in the creative work it produces.

Perceptions of Creative Quality

As people become more familiar or exposed to a particular social referent, they can update their general impressions of the referent's ability (Smith & Arnkelsson, 2000), changing the consequences this target has for social comparative self-judgments. For example, our supplemental data (described in the prior section) highlight how gen-AI can be perceived overall as better or worse at generating creative content, shifting the levels of self-confidence that results from the associated social comparison processes. However, orthogonal to general perceptions of the target, perceptions of a specific piece of work's quality can also influence self-perceptions through a comparative process. As we observed in Study 4, the perceived quality of the story influenced perceivers' self-confidence (e.g., people tended to report greater self-confidence after viewing lower quality work) separately from the effect of the work's ostensible source. So, might simply perceiving creative content as originating from gen-AI lead to lower quality perceptions of the specific work, and is this what primarily influences people's self-confidence?

Across our studies, we additionally collected measures of the work's quality (e.g., how funny is this joke? how enjoyable is this story?), allowing us to test whether (a) labeling the content as created by gen-AI versus human affected perceptions of the work's quality, and (b) whether this played an equivalent or greater role relative to perceptions of the author's general abilities in mediating the self-confidence effect. All details can be found in the Supplemental Web Appendix, but in short, we found (a) a less consistent effect of labeling on perception of the work's quality (vs. perceptions of the source's ability) and (b) that perceptions of quality (vs. the source's ability) less commonly mediated the confidence effects. This suggests that perceiving a source as gen-AI (vs. human) can lead to perceptions of lower quality work (in turn enhancing people's own confidence to create similar creative work); however, the generalized impressions people hold about gen-AI's creative abilities more reliably drove the observed self-confidence effect. Admittedly, there might be certain domains where people's self-judgments are more strongly anchored on the quality of the work (vs. perceptions of the author's ability). For example, when the quality of the work is confirmed by others' views (e.g., a person sees that multiple people find a joke funny), they might be more inclined to use this as their input for social comparison, rather than perceptions of the source. We leave this to future research.

Future Directions and Limitations

Given the prevalence of gen-AI content online (McKinsey & Company, 2024), we believe there is much more work that can be

Table 1
Limitations of and Future Directions

Limitation	Explanation of limitation	Future direction
Non-Western samples	Our sample all came from online convenience samples in either the United States or the United Kingdom.	Studying these effects on individuals who are less familiar with gen-AI or those from collectivistic cultures (i.e., where social comparison processes might operate differently) could offer interesting boundary conditions for the effects.
Comparative self-confidence measure	Across our primary studies, we relied on comparative (vs. absolute) measures of self-confidence. Although we tested the effects with absolute measures in Study 3 and the supplementary studies, the strength of the effect or its correspondence with other effects might hinge on the type of measure employed.	Using absolute measures of self-confidence and contrasting it to comparative measures could inform whether one type of measure is more sensitive to the effect. It could also be helpful to determine whether one type of measure better corresponds to different types of downstream behaviors.
Alternative processes	We provided evidence for a social comparison process, focusing primarily on gen-AI as a social referent. We additionally tested how perceptions of the quality of the work could matter; however, there are other alternative processes we did not explore. Additionally, our primary support for our proposed process came through statistical mediation.	Even robust social comparison effects can be multiply influenced, and future research would benefit from better understanding how other perceptions about gen-AI, the human-peer comparison, or the quality of the work influence the effects. Additionally, employing other process-by-moderation designs could provide greater confidence in the proposed social comparison process.
Conditional boundary conditions	The present research provided one context-based boundary condition (i.e., the domain of the output); however, there are likely other contexts and individual differences that would moderate the findings and their impact.	Understanding the variety of domains in which gen-AI serves as a higher or lower social referent (e.g., possibly at legal matters vs. therapy, respectively) will offer important insight on when the enhanced self-confidence effect is more or less likely to emerge.
Behavioral consequences	The current research showed that people's enhanced creative self-confidence could increase people's willingness to engage in the same activity; however, only one study had them actually complete a relevant activity. Moreover, there are likely other behaviors that this self-confidence effect might or might not affect.	Although self-confidence has been shown to affect all different types of behaviors, a better understanding of the taxonomy of what would be affected by the enhanced self-confidence in this work could help both theory and practice in understanding the limits of the effect and its real-world application.

Note. gen-AI = generative artificial intelligence.

done. Foremost, is to understand how long these self-confidence effects persist. Although even transitory influences on self-confidence can have significant ramifications (Andrade, 2015; Greenacre et al., 2014; Meisha & Al-Dabbagh, 2021), understanding the degree to which the effects observed in this research persist in the long term offers its own insights. We also believe it would be worth testing different ecologically useful interventions that this self-confidence effect could aid. For example, educators might provide explicitly gen-AI created examples of various creative content (e.g., a short story, one's personal narrative for college) to enhance self-confidence and promote taking "the first step" in creative tasks that students would have otherwise lacked the self-confidence to pursue. In considering these future directions, it is important to consider the limitations of the present research (see Table 1). Even with these caveats, though, people reliably reported greater creative self-confidence following exposure to gen-AI labeled work, offering a new line of questions as to how gen-AI can and will serve as a social referent for other social comparison processes and effects.

References

Andrade, V. (2015). The importance of models of self-confidence in sports performance. *Journal of Social and Psychological Sciences*, *8*(1), 1–8.

Baek, T. H., Kim, H. J., & Kim, J. (2025). AI-generated recommendations: Roles of language style, perceived AI human-likeness, and recommendation agent. *International Journal of Hospitality Management*, *126*, Article 104106. https://doi.org/10.1016/j.ijhm.2025.104106

Balmer, A., & Murcott, A. (2020). How to get past writer's block. *The craft of writing in sociology* (pp. 152–153). Manchester University Press. https://doi.org/10.7765/9781526153364.00017

Bandura, A. (1997). *Self-efficacy: The exercise of control* (Vol. 604). Freeman.

Beghetto, R. A. (2006). Creative self-efficacy: Correlates in middle and secondary students. *Creativity Research Journal*, *18*(4), 447–457. https://doi.org/10.1207/s15326934crj1804_4

Bojić, L., Stojković, I., & Jolić Marjanović, Z. (2024). Signs of consciousness in AI: Can GPT-3 tell how smart it really is? *Humanities and Social Sciences Communications*, *11*(1), Article 1631. https://doi.org/10.1057/s41599-024-04154-3

Burson, K. A. (2007). Consumer-product skill matching: The effects of difficulty on relative self-assessment and choice. *Journal of Consumer Research*, *34*(1), 104–110. https://doi.org/10.1086/513051

Carmeli, A., & Schaubroeck, J. (2007). The influence of leaders' and other referents' normative expectations on individual involvement in creative work. *The Leadership Quarterly*, *18*(1), 35–48. https://doi.org/10.1016/j.leaqua.2006.11.001

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Chatterjee, A. (2014). *The aesthetic brain: How we evolved to desire beauty and enjoy art*. Oxford University Press.

Cheek, N. N., Blackman, S. F., & Pronin, E. (2021). Seeing the subjective as objective: People perceive the taste of those they disagree with as biased and wrong. *Journal of Behavioral Decision Making*, *34*(2), 167–182. https://doi.org/10.1002/bdm.2201

Collins, R. L. (2000). Among the better ones. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison. The Springer series in social clinical psychology* (pp. 159–171). Springer. https://doi.org/10.1007/978-1-4615-4237-7_9

De Vito, E. L. (2023). Artificial intelligence and chatGPT. Would you read an artificial author? *Medicina*, *83*(2), 329–332.

Denrell, J., & Liu, C. (2012). Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9331–9336. https://doi.org/10.1073/pnas.1116048109

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, *384*(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests—An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, *75*, 95–102. https://doi.org/10.1016/j.jesp.2017.11.008

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, *138*(2), 296–321. https://doi.org/10.1037/a0026556

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, *144*(2), 177–197. https://doi.org/10.1037/bul0000127

Ginns, P., Martin, A. J., Freebody, K., Anderson, M., & O'Connor, P. (2023). Creative self-beliefs among children and adolescents. *Australian Journal of Education*, *67*(2), 181–195. https://doi.org/10.1177/00049441231177097

Goffin, R. D., & Olson, J. M. (2011). Is it all relative? Comparative judgments and the possible improvement of self-ratings and ratings of others. *Perspectives on Psychological Science*, *6*(1), 48–60. https://doi.org/10.1177/1745691610393521

Greenacre, L., Tung, N. M., & Chapman, T. (2014). Self confidence, and the ability to influence. *Academy of Marketing Studies Journal*, *18*(2), 169–180.

Harreis, H., Koullias, T., Roberts, R., & Te, K. (2023). *Generative AI: Unlocking the future of fashion*. McKinsey & Company.

Haslam, S. A., Adarves-Yorno, I., Postmes, T., & Jans, L. (2013). The collective origins of valued originality: A social identity approach to creativity. *Personality and Social Psychology Review*, *17*(4), 384–401. https://doi.org/10.1177/1088868313498001

Hayes, A. F. (2022). *Counterfactual/potential outcomes "causal mediation analysis" with treatment by mediator interaction using PROCESS*. Canadian Centre for Research Analysis and Methods Technical Report.

Horwood, S., & Anglim, J. (2019). Problematic smartphone usage and subjective and psychological well-being. *Computers in Human Behavior*, *97*, 44–50. https://doi.org/10.1016/j.chb.2019.02.028

Ipeirotis, P. G. (2010). Analyzing the Amazon Mechanical Turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, *17*(2), 16–21. https://doi.org/10.1145/1869086.1869094

Karwowski, M. (2016). The dynamics of creative self concept: Changes and reciprocal relations between creative self-efficacy and creative personal identity. *Creativity Research Journal*, *28*(1), 99–104. https://doi.org/10.1080/10400419.2016.1125254

Karwowski, M., & Barbot, B. (2016). Creative self-beliefs: Their nature, development, and correlates. In J. C. Kaufman & J. Baer (Eds.), *Creativity and reason in cognitive development* (2nd ed., pp. 302–326). Cambridge University Press. https://doi.org/10.1017/CBO9781139941969.016

Korteling, J. E. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human-versus artificial intelligence. *Frontiers in Artificial Intelligence*, *4*, Article 622364. https://doi.org/10.3389/frai.2021.622364

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, *53*(9), 1017–1031. https://doi.org/10.1037/0003-066X.53.9.1017

Latané, B. (1966). Studies in social comparison—Introduction and overview. *Journal of Experimental Social Psychology*, *1*(Suppl. 1), 1–5. https://doi.org/10.1016/0022-1031(66)90060-6

Lee, Y. K., Chang, C. T., Lin, Y., & Cheng, Z. H. (2014). The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress. *Computers in Human Behavior*, *31*, 373–383. https://doi.org/10.1016/j.chb.2013.10.047

Lemons, G. (2010). Bar drinks, rugas, and gay pride parades: Is creative behavior a function of creative self-efficacy? *Creativity Research Journal*, *22*(2), 151–161. https://doi.org/10.1080/10400419.2010.481502

Lily, A. E., Ismail, A. F., Abunaser, F. M., Al-Lami, F., & Abdullatif, A. K. A. (2023). ChatGPT and the rise of semi-humans. *Humanities and Social Sciences Communications*, *10*(1), Article 626. https://doi.org/10.1057/s41599-023-02154-3

Marriott, H. R., & Pitardi, V. (2024). One is the loneliest number… Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction. *Psychology and Marketing*, *41*(1), 86–101. https://doi.org/10.1002/mar.21899

Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports*, *14*(1), Article 4692. https://doi.org/10.1038/s41598-024-53755-0

McKenna, K. Y., & Bargh, J. A. (2014). Plan 9 from cyberspace: The implications of the Internet for personality and social psychology. In M. B. Brewer, D. A. Kenny, & J. K. Norem (Eds.), *Personality and social psychology at the interface* (pp. 57–75). Psychology Press.

McKinsey & Company. (2024, May 30). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Meisha, D. E., & Al-Dabbagh, R. A. (2021). Self-confidence as a predictor of senior dental student academic success. *Journal of Dental Education*, *85*(9), 1497–1503. https://doi.org/10.1002/jdd.12617

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*(3), 221–237. https://doi.org/10.1177/1745691612441215

Morewedge, C. K. (2022). Preference for human, not algorithm aversion. *Trends in Cognitive Sciences*, *26*(10), 824–826. https://doi.org/10.1016/j.tics.2022.07.007

Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology*, *16*(1), 148–156. https://doi.org/10.1037/h0029862

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*(3), 472–489. https://doi.org/10.1037/0033-295X.110.3.472

Perplexity. (2024, June 11). *Write me a joke about Lexapro* [Computer software]. Perplexity AI. https://www.perplexity.ai

Puente-Díaz, R., & Cavazos-Arroyo, J. (2017). Creative self-efficacy: The influence of affective states and social persuasion as antecedents and imagination and divergent thinking as consequences. *Creativity Research Journal*, *29*(3), 304–312. https://doi.org/10.1080/10400419.2017.1360067

Reeder, G. D., Hesson-McInnis, M., Krohse, J. O., & Scialabba, E. A. (2001). Inferences about effort and ability. *Personality and Social Psychology Bulletin*, *27*(9), 1225–1235. https://doi.org/10.1177/0146167201279014

Royston, R., & Reiter-Palmon, R. (2019). Creative self-efficacy as mediator between creative mindsets and creative problem-solving. *The Journal of Creative Behavior*, *53*(4), 472–481. https://doi.org/10.1002/jocb.226

Santoro, E., & Monin, B. (2023). The AI effect: People rate distinctively human attributes as more essential to being human after learning about artificial intelligence advances. *Journal of Experimental Social Psychology*, *107*, Article 104464. https://doi.org/10.1016/j.jesp.2023.104464

Schweitzer, S., Dobson, K. S., & Waytz, A. (2024). Political bot bias in the perception of online discourse. *Social Psychological and Personality Science*, *15*(2), 234–244. https://doi.org/10.1177/19485506231156020

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Siev, J. J., & Teeny, J. D. (2024). Personal misconduct elicits harsher professional consequences for artists (vs. scientists): A moral-decoupling process. *Psychological Science*, *35*(1), 82–92. https://doi.org/10.1177/09567976231214739

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-positive citations. *Perspectives on Psychological Science*, *13*(2), 255–259. https://doi.org/10.1177/1745691617698146

Smith, W. P., & Arnkelsson, G. B. (2000). Stability of related attributes and the inference of ability through social comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison. The Springer series in social clinical psychology* (pp. 45–66). Springer. https://doi.org/10.1007/978-1-4615-4237-7_3

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845–851. https://doi.org/10.1037/0022-3514.89.6.845

Stemler, S. E., & Kaufman, J. C. (2020). Are creative people better than others at recognizing creative work? *Thinking Skills and Creativity*, *38*, Article 100727. https://doi.org/10.1016/j.tsc.2020.100727

Sufi, F. (2024). Generative pre-trained transformer (GPT) in research: A systematic review on data augmentation. *Information*, *15*(2), Article 99. https://doi.org/10.3390/info15020099

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science*, *11*(5), 159–163. https://doi.org/10.1111/1467-8721.00191

Sun, L., Yuan, Y., Yao, Y., Li, Y., Zhang, H., Xie, X., Wang, X., Luo, F., & Stillwell, D. (2024). *Large language models show both individual and collective creativity comparable to humans*. arXiv preprint. https://doi.org/10.48550/arXiv.2412.03151

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal*, *45*(6), 1137–1148. https://doi.org/10.2307/3069429

Tierney, P., & Farmer, S. M. (2011). Creative self-efficacy development and creative performance over time. *Journal of Applied Psychology*, *96*(2), 277–293. https://doi.org/10.1037/a0020952

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Unkelbach, C., Alves, H., Baldwin, M., Crusius, J., Diel, K., Galinsky, A. D., Gast, A., Hofmann, W., Imhoff, R., Genschow, O., Lammers, J., Pauels, E., Schneider, I., Topolinski, S., Westfal, M., & Mussweiler, T. (2023). Relativity in social cognition: Basic processes and novel applications of social comparisons. *European Review of Social Psychology*, *34*(2), 387–440. https://doi.org/10.1080/10463283.2022.2161043

Walumbwa, F. O., Christensen-Salem, A., Hsu, I. C., & Misati, E. (2018). Creative self-efficacy and creative performance: Understanding the underlying mechanisms. *Academy of Management Proceedings*, *2018*(1), Article 10208. https://doi.org/10.5465/AMBPP.2018.75

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*(4), 393–407. https://doi.org/10.1037/gpr0000056

Ward, A. F. (2021). People mistake the internet's knowledge for their own. *Proceedings of the National Academy of Sciences*, *118*(43), Article e2105061118. https://doi.org/10.1073/pnas.2105061118

Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research*, *2*(2), 140–154. https://doi.org/10.1086/691462

Wegner, D. M., & Ward, A. F. (2013). How Google is changing your brain. *Scientific American*, *309*(6), 58–61. https://doi.org/10.1038/scientificamerican1213-58

Wheeler, L. (1991). A brief history of social comparison theory. In J. Suls & T. A. Wills (Eds.), *Social comparison* (pp. 3–21). Routledge.

Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin*, *106*(2), 231–248. https://doi.org/10.1037/0033-2909.106.2.231

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi.org/10.1177/1745691613518075

Received October 23, 2024
Revision received February 27, 2025
Accepted March 18, 2025 ■

Web Appendix

Pilot Data

To get a sense of how often people are exposed to gen-AI content, we recruited a sample of social media users on *Connect*, where we restricted our sample to those who use Facebook, Reddit, and/or Instagram (N = 150, M_{age} = 39.16, SD = 11.08; male = 52.7%, female = 46.0%, non-binary = 0.7%, prefer not to answer = 0.7%).

The pilot began by informing participants that we were interested in their experience and exposure to content on social media. Participants then reported how often on social media they see "visual art" (which we defined: "Visual art can include drawings, illustrations, or photography"). This was assessed on a scale ranging from 1 = None at all to 10 = A great deal. Next, participants reported on the percentage of visual art they see on social media that they believe is "created by generative artificial intelligence (AI)." (which we defined: "Generative AI are machine-algorithms like ChatGPT, DALL-E, Gemini, which can create visual art similar to what humans produce."). This was assessed on a scale ranging from 0% to 100%. After answering these questions about visual art, they answered the same series of questions about the "humorous writing" they see on social media (which we defined: "Humorous writing can include jokes, comics, or witty commentary"). Participants reported on the amount of humorous content they see and the percentage of which they believed was created by generative AI. See the table below for the means and standard deviation for both measures.

	Mean (SD)
Overall Visual Art	7.04 (2.20)
Percentage of Visual Art from Gen-AI	42.40% (22.99)
Overall Humorous Content	6.97 (2.04)
Percentage of Humorous Content from Gen-AI	29.33% (23.37)

Absolute Measures of Self-Confidence

As noted in the manuscript, Study 3 explicitly employed an absolute measure of self-confidence; however, it used language that did not reference "confidence" specifically. Thus, to provide better evidence as to whether measures assessing the absolute "confidence" and "certainty" in one's ability to complete a creative task would be affected by the source of the content (gen-AI vs. human), in two of our supplemental studies, we used absolute versions of our primary, relative confidence measure. These studies were replications of our two supplemental studies (reported in detail in a later section), that focused on explicitly using absolute measures of self-confidence. More specifically, in two separate studies we asked participants "How confident are you that you can come up with a short story?" and "How certain are you that you can come up with a short story?" (1 = Not at all, 7 = Extremely).

In the first of these studies (N = 797), we observed a significant effect on self-confidence, where participants exposed to creative content ostensibly written by gen-AI (vs. a human) resulted in greater self-confidence, now, on an absolute measure (M_{AI} = 4.13, SD = 1.71 vs. M_{human} = 3.68, SD = 1.70; F(1,793) = 13.84, p < .001). In the second study (N = 800), the same was true (M_{AI} = 4.53, SD = 1.76 vs. M_{human} = 4.20, SD = 1.76; F(1,796) = 7.15, p = .008).

Altogether, these studies indicate that the phenomenon we observed in our primary studies was not an artifact of the comparative version of the measure we used. Nonetheless, future research could benefit from determining whether there are instances when one type of measure is more sensitive to detecting changes in self-confidence given the surrounding context.

Analysis with Inattentive Participants

In the primary text, we report results from our studies with all participants included. We did this to minimize post-hoc exclusion decisions and maximize our studies' power. This approach also allows for detecting effects that are more robust to "noise" in the data. However, it can nonetheless be valuable to determine how the effect emerges for people attentive to the manipulation as well as the study's context. Additionally, using exclusion criteria can help to remove error variance, increasing the study's power and ensuring that any effects are not due to random chance. To this effect, we re-analyzed our data using the following exclusion criteria in consideration of the measures we had available:

- *Participants with duplicate IP addresses*. Because participants who take the study twice might either be trying to scam the study for money, or one of their responses might be less reflective of their true psychology, we take the most conservative approach and exclude anyone with a duplicate IP address. Notably, we only observed this in a subset of studies, and for those where it did occur, there was only ever one pair of duplicate IP addresses.
- *Impossibly short completion times*. Although there will be variation in the length of time it takes a participant to complete a study, some completion times are so fast that it would be humanly impossible to read all of the content/instructions, while providing true responses to the various measures. At first blush, one might suggest taking an extreme outlier approach, where, for example, people who complete a study 3 standard deviations (SD) faster than the average completion time would be excluded. However, because completion time has a lower bound but no upper bound, this means that many times a single SD exceeds 0 (i.e., variation of the upper bound results in an SD that exceeds the limits of the lower bound). To this effect, we more simply excluded anyone who completed any of the studies in 20 seconds or fewer, which would be an impossibly short amount of time to read the content and respond to the measures with any sort of experimental fidelity.

The results on our primary self-confidence measure from this analysis can be found below. In short, all of the self-confidence effects reported in the primary analysis are replicated in the reduced samples:

	Reduced Sample Size	Gen-AI Condition	Human Condition	p-value
Study 1A	398	3.70	3.24	.010
Study 1B	391	4.08	3.00	< .001
Study 1C	397	3.75	2.90	< .001
Study 2	395	4.99	3.88	< .001
Study 3	399	3.77	3.06	< .001
Study 4	783	3.59	3.16	< .001
Study 5	799	4.20	3.70	< .001

Zero-Order Correlations between Primary Measures

Given the potential similarities in the measures we used to assess our primary DV (self-confidence), mechanism (perceptions of the target's relevant ability), and control variable (perceptions of the quality's work), we report the zero-order correlations between all of these measures. To summarize these findings, we observe empirical relationships in line with our theoretical assumptions. First, we find that the highest correlations are consistently the two items we use to form an index (e.g., our two self-confidence measures). Second, the next highest correlation (which nonetheless vary from low to medium-high) is the positive relationship between perceptions of the work's quality and generalized perceptions of the individual. For example, perceptions that the work is higher quality is associated with perceptions that the author is a more skilled creator. Third, we see stronger associations between perceptions of the author's ability (our proposed mechanism) and self-confidence than we do between perceptions of the work's quality (our control measure) and self-confidence. All in all, these results help to support the notion that our measures are capturing different constructs that reflect the theoretically proposed relationships in the primary work.

Study 1A

Correlations

		Q12 How funny is the joke about Lexapro?	Q30 To what extent do you think you could have come up with a better joke?	Q14 To what extent does the joke writer have a sense of humor?
Q12 How funny is the joke about Lexapro?	Pearson Correlation	1	-.176**	.661**
	Sig. (2-tailed)		<.001	<.001
	N	400	400	400
Q30 To what extent do you think you could have come up with a better joke?	Pearson Correlation	-.176**	1	-.193**
	Sig. (2-tailed)	<.001		<.001
	N	400	400	400
Q14 To what extent does the joke writer have a sense of humor?	Pearson Correlation	.661**	-.193**	1
	Sig. (2-tailed)	<.001	<.001	
	N	400	400	400

**. Correlation is significant at the 0.01 level (2-tailed).

Study 1B

Correlations

		Q109 How much did you enjoy this drawing?	Q37 To what extent is the creator of this piece a good drawer?	Q36 How confident are you that you could have come up with a better drawing symbolizing protection?	Q68 How certain are you that you could have come up with a better drawing symbolizing protection?
Q109 How much did you enjoy this drawing?	Pearson Correlation	1	.498**	-.342**	-.285**
	Sig. (2-tailed)		<.001	<.001	<.001
	N	399	399	399	399
Q37 To what extent is the creator of this piece a good drawer?	Pearson Correlation	.498**	1	-.357**	-.330**
	Sig. (2-tailed)	<.001		<.001	<.001
	N	399	399	399	399
Q36 How confident are you that you could have come up with a better drawing symbolizing protection?	Pearson Correlation	-.342**	-.357**	1	.890**
	Sig. (2-tailed)	<.001	<.001		<.001
	N	399	399	399	399
Q68 How certain are you that you could have come up with a better drawing symbolizing protection?	Pearson Correlation	-.285**	-.330**	.890**	1
	Sig. (2-tailed)	<.001	<.001	<.001	
	N	399	399	399	399

**. Correlation is significant at the 0.01 level (2-tailed).

Study 1C

Correlations

		enj How much did you enjoy this poem?	write To what extent is the writer of this poem good at writing poetry?	con1 How confident are you that you could have come up with a better nature poem?	con2 How certain are you that you could have come up with a better nature poem?
enj How much did you enjoy this poem?	Pearson Correlation	1	.790**	-.228**	-.215**
	Sig. (2-tailed)		<.001	<.001	<.001
	N	400	399	400	400
write To what extent is the writer of this poem good at writing poetry?	Pearson Correlation	.790**	1	-.320**	-.301**
	Sig. (2-tailed)	<.001		<.001	<.001
	N	399	399	399	399
con1 How confident are you that you could have come up with a better nature poem?	Pearson Correlation	-.228**	-.320**	1	.901**
	Sig. (2-tailed)	<.001	<.001		<.001
	N	400	399	400	400
con2 How certain are you that you could have come up with a better nature poem?	Pearson Correlation	-.215**	-.301**	.901**	1
	Sig. (2-tailed)	<.001	<.001	<.001	
	N	400	399	400	400

**. Correlation is significant at the 0.01 level (2-tailed).

Study 2

Correlations

		enjoy How much did you enjoy this story?	interesting How interesting was this story?	writer To what extent is the writer of the story a good writer?	confident How confident are you that you could have come up with a better story?	certain How certain are you that you could have come up with a better story?	will As part of this study, would you like to write a story using the story spine?
enjoy How much did you enjoy this story?	Pearson Correlation	1	.849**	.718**	-.249**	-.253**	.069
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	.169
	N	400	400	400	400	400	400
interesting How interesting was this story?	Pearson Correlation	.849**	1	.735**	-.270**	-.276**	.051
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	.307
	N	400	400	400	400	400	400
writer To what extent is the writer of the story a good writer?	Pearson Correlation	.718**	.735**	1	-.396**	-.402**	-.033
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	.516
	N	400	400	400	400	400	400
confident How confident are you that you could have come up with a better story?	Pearson Correlation	-.249**	-.270**	-.396**	1	.932**	.214**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001
	N	400	400	400	400	400	400
certain How certain are you that you could have come up with a better story?	Pearson Correlation	-.253**	-.276**	-.402**	.932**	1	.197**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		<.001
	N	400	400	400	400	400	400
will As part of this study, would you like to write a story using the story spine?	Pearson Correlation	.069	.051	-.033	.214**	.197**	1
	Sig. (2-tailed)	.169	.307	.516	<.001	<.001	
	N	400	400	400	400	400	400

**. Correlation is significant at the 0.01 level (2-tailed).

Study 3

Correlations

		Q109 How funny is this caption?	writer To what extent does the caption writer have a sense of humor?	Q36 How confident are you that you could have come up with a better caption?	Q40 How certain are you that you could have come up with a better caption?	Q41_1 Please rate the extent your caption: [QID1717658856-ChoiceTextEntryValue] - Is funny	Q41_2 Please rate the extent your caption: [QID1717658856-ChoiceTextEntryValue] - Made you laugh
Q109 How funny is this caption?	Pearson Correlation	1	.752**	-.097	-.072	.269**	.279**
	Sig. (2-tailed)		<.001	.052	.151	<.001	<.001
	N	401	401	401	401	401	401
writer To what extent does the caption writer have a sense of humor?	Pearson Correlation	.752**	1	-.167**	-.163**	.152**	.173**
	Sig. (2-tailed)	<.001		<.001	.001	.002	<.001
	N	401	401	401	401	401	401
Q36 How confident are you that you could have come up with a better caption?	Pearson Correlation	-.097	-.167**	1	.919**	.466**	.473**
	Sig. (2-tailed)	.052	<.001		<.001	<.001	<.001
	N	401	401	401	401	401	401
Q40 How certain are you that you could have come up with a better caption?	Pearson Correlation	-.072	-.163**	.919**	1	.503**	.516**
	Sig. (2-tailed)	.151	.001	<.001		<.001	<.001
	N	401	401	401	401	401	401
Q41_1 Please rate the extent your caption: [QID1717658856-ChoiceTextEntryValue] - Is funny	Pearson Correlation	.269**	.152**	.466**	.503**	1	.872**
	Sig. (2-tailed)	<.001	.002	<.001	<.001		<.001
	N	401	401	401	401	401	401
Q41_2 Please rate the extent your caption: [QID1717658856-ChoiceTextEntryValue] - Made you laugh	Pearson Correlation	.279**	.173**	.473**	.516**	.872**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	
	N	401	401	401	401	401	401

**. Correlation is significant at the 0.01 level (2-tailed).

Study 4

Correlations

		Q51 How funny is this caption?	writer To what extent does the caption writer have a sense of humor?	Q36 How confident are you that you could have come up with a better caption?	Q40 How certain are you that you could have come up with a better caption?
Q51 How funny is this caption?	Pearson Correlation	1	.877**	-.311**	-.321**
	Sig. (2-tailed)		<.001	<.001	<.001
	N	802	802	802	802
writer To what extent does the caption writer have a sense of humor?	Pearson Correlation	.877**	1	-.321**	-.329**
	Sig. (2-tailed)	<.001		<.001	<.001
	N	802	802	802	802
Q36 How confident are you that you could have come up with a better caption?	Pearson Correlation	-.311**	-.321**	1	.923**
	Sig. (2-tailed)	<.001	<.001		<.001
	N	802	802	802	802
Q40 How certain are you that you could have come up with a better caption?	Pearson Correlation	-.321**	-.329**	.923**	1
	Sig. (2-tailed)	<.001	<.001	<.001	
	N	802	802	802	802

**. Correlation is significant at the 0.01 level (2-tailed).

Study 5

Correlations

		qual1	qual2	conf1	conf2	writer
qual1	Pearson Correlation	1	.839**	-.353**	-.362**	.711**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001
	N	802	802	802	802	802
qual2	Pearson Correlation	.839**	1	-.447**	-.461**	.787**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001
	N	802	802	802	802	802
conf1	Pearson Correlation	-.353**	-.447**	1	.939**	-.571**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001
	N	802	802	802	802	802
conf2	Pearson Correlation	-.362**	-.461**	.939**	1	-.577**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001
	N	802	802	802	802	802
writer	Pearson Correlation	.711**	.787**	-.571**	-.577**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	
	N	802	802	802	802	802

**. Correlation is significant at the 0.01 level (2-tailed).

General Discussion Studies: Emotional Depth and Authenticity

As described in the General Discussion of the primary manuscript, we ran two additional studies to test the extent to which people were able to update their naïve theory about gen-AI's creativity. To that effect, we used paradigms similar to our primary studies (e.g., Study 3), while adding an additional factor that described the source of the creative work (i.e., a fellow Prolific worker or gen-AI) as varying on one of the two traits fundamental to creativity: emotional depth and authenticity. If people are generally unable to update their perceptions of gen-AI's creative ability, we should see a similar effect on (no difference between) participants' creative self-confidence in the low (or baseline) and high trait conditions. In contrast, we should see a change in self-confidence for the human conditions (e.g., as humans are seen as having more emotional depth, they should be seen as a higher social referent, lowering participants' creative self-confidence). Ultimately, we observed that both the source manipulation (gen-AI vs. human) and the creativity manipulation (baseline vs. high) or (low vs. high) affected self-confidence ratings and perceptions of the target without interacting. This suggests that perceptions of gen-AI as a social referent on this dimension can update similarly to how perceptions of other humans can update, influencing perceivers' creative self-confidence similarly.

Emotional Depth

In the emotional depth study, we recruited 800 participants (M_{age} = 37.83, SD = 13.38; male = 47.4%, female = 50.8%, non-binary = 1.8%) who were randomly assigned to a 2 (source: gen-AI vs. human) x 2 (emotional depth: control vs. high) between-subjects design.

Participants followed a paradigm very similar to Study 2. That is, they began by learning that they would read a short story by either a gen-AI or Prolific participant. Unlike the prior studies, however, participants were also provided with additional information about the source of the short story: their previous work either scored "high" on emotional depth or no further information was provided. This was then repeated to them before they read a short story about a first-person account of falling in love. Unlike Study 2 in the primary manuscript, this short story was not structured in the form of a story spine and instead read as a piece of flash fiction (roughly 100 words). The story described a sunny afternoon, where the narrator meets someone and has a wonderful time. After reading the story, participants reported their self-confidence on two measures used throughout this work ("How confident are you that you can come up with a better short story about the first time you fell in love?" and "How certain are you that you can come up with a better short story about the first time you fell in love" 1 = Not at all, 7 = Extremely; r = .93).

We conducted a 2 (source: gen-AI vs. human) x 2 (emotional depth: low vs. high) ANOVA on the self-confidence measure. From this, we observed a significant main effect of the source, replicating our earlier work: believing the short story was written by gen-AI (vs. a human) resulted in greater creative self-confidence (M_{AI} = 4.60, SD = 1.82 vs. M_{human} = 3.85, SD = 1.79; F(1,796) = 34.55, p < .001). We also observed a main effect of the emotional depth manipulation, where the author described as being higher in emotional depth resulted in lower self-confidence (M_{high} = 4.07, SD = 1.82 vs. $M_{control}$ = 4.38, SD = 1.86; F(1,796) = 5.70, p = .017) There was no significant interaction (p = .103).

Authenticity

In the authenticity study, we recruited 800 participants (M_{age} = 40.97, SD = 13.38; male = 44.8%, female = 53.1%, non-binary = 1.3%, prefer not to say = 0.9%) who were randomly assigned to a 2 (source: gen-AI vs. human) x 2 (authenticity: low vs. high) between-subjects design.

This study was similar to the emotional depth study described above with three primary changes. First, rather than describing the author of the short story to vary in terms of their emotional depth, we varied their perceived authenticity, using a similar wording manipulation. Second, rather than testing high versus a control condition, we explicitly described the author to be "high" or "low" in terms of authenticity. Third, and in order to eliminate any effect that the perception of the work and/or the topic of the work might have on the effect, we asked participants to report how much confidence/certainty they have in their ability to come up with a short story (on the same measures as the prior study; r = .91) without actually reading a short story. That is, we told participants that they would be reading a short story authored in line with one of the four possible conditions, allowing us to see how a pure test of the generalized perception of the author's abilities at writing would affect participants' self-confidence.

Conducting another 2 (source: gen-AI vs. human) x 2 (authenticity: low vs. high) ANOVA on the self-confidence measure, we again observed a significant main effect: believing the short story was written by gen-AI (vs. a human) resulted in greater creative self-confidence (M_{AI} = 3.98, SD = 1.66 vs. M_{human} = 3.70, SD = 1.58; F(1,796) = 6.21, p = .013). We also observed a main effect of the emotional depth manipulation, where the author described as being higher in emotional depth resulted in lower self-confidence (M_{high} = 3.46, SD = 1.69 vs. $M_{control}$ = 4.22, SD = 1.47; F(1,796) = 47.32, p < .001) There was no significant interaction (p = .146).

Understanding the Role of the Content's Perceived Quality

In the focal manuscript, we provide consistent evidence that perceptions of generative artificial intelligence (gen-AI) as a lower social referent (relevant to a human peer) at creative abilities played a focal role in the observed effects. This social comparison process was supported both mediationally as well as through experimental moderation. However, another plausible contributor to the effect could be the perceived quality of the work itself. That is, might labeling creative content as created by gen-AI (vs. human) result in perceiving the work itself as lower quality? If so, people might consequently report enhanced creative self-confidence after observing gen-AI work, because they perceive the work itself to be below their own capabilities.

In our work, specifically, Study 4 of the primary text, we found that the social referent process could enhance people's creative self-confidence above the effect of people's perceived quality of the work. That is, we showed that people's creative self-confidence was anchored on their perception of gen-AI's overall ability at the task as the strength of the effect was not moderated by the quality of the specific creative output. We suspect this might have occurred because the evaluation of creative work is subjective, meaning there is greater latitude for individuals to rely on their perceptions of the source (gen-AI vs. human), rather than on an "objective assessment" of the content itself. Still, this is not to say that the objective quality of the work itself cannot influence creative self-confidence judgments (e.g., observing extremely low quality creative content from an ostensibly high social referent could enhance confidence judgments). Indeed, we did observe a main effect in Study 4, where people reported lower creative self-confidence after being exposed to the higher (vs. lower) quality work. However, presuming a normal range of quality for the output, we would suspect that individuals are likely to feel more confident in their own abilities when comparing themselves to gen-AI (a stable, lower social referent), irrespective of the creative content's higher or lower quality.

Now, as we have been describing, the source for social comparison-based self-judgments can indeed stem from the quality of the content itself, too. To reiterate, after exposure to lower (higher) quality output, people tend to report greater (reduced) confidence in their own relevant abilities. Thus, it could be that simply perceiving creative content as originating from gen-AI (vs. humans), could lead to lower perceptions of the work's quality, which could amplify people's own confidence in their abilities. Because we collected measures of the work's quality across our studies (e.g., how funny is this joke? how enjoyable is this story?), we can test such questions. Namely, we can test whether (a) labeling the content as created by gen-AI versus human affected perceptions of the work's quality, and (b) whether this played an equivalent or greater role in mediating the effect of the content's source (gen-AI vs. human) on self-confidence relative to perceptions of the author's innate abilities. The results for both of these questions are presented in Table 1, where, in short, we find (a) a less consistent effect of labeling on perception of the work's quality and (b) that perceptions of quality less commonly mediate the confidence effects.

Across our various datasets, although perceptions of the quality's work were often affected by whether the creative content was ostensibly created by gen-AI versus humans, it less frequently played a significant role in explaining the creative self-confidence effect. This suggests that perceiving gen-AI (vs. human) work can lead to perceptions that it is of lower quality (in turn enhancing people's own confidence to create similar creative work); however, it is less reliable than the process we focused on. People simply perceive gen-AI (vs. humans) to be less capable of generating creative work, and therefore exposure to the work it produces leads to enhanced confidence in peoples' own relevant abilities. Again, this takeaway is further supported

by the results of Study 4, where we found no difference in the size of the confidence effects after participants observed objectively worse and better content. Still, there could be important reciprocal effects in this line of work that future research could address. For example, expecting gen-AI to be worse at creating creative content could lead to greater creative self-confidence effects when the quality of the work is very low, while it could lead to weaker effects when the quality of the work is exceptionally (rather than just moderately) high. Additionally, there might be certain domains where people's self-judgments are more strongly anchored on the quality of the work, relative to their perceptions of the author's ability. For example, when the quality of the work is confirmed by others' views (e.g., a person sees that multiple people find a joke funny), they might be more inclined to use this as their source for social comparison.

Table 1. The Role of Perceived Quality: Conditional Effects and Parallel Mediation

	Content Quality Perceptions			Parallel Mediation (boldened CIs indicate significant indirect effects in joint mediation)	
	Gen-AI M (SD)	Human M (SD)	p-value	Content Quality	Social Referent
Study 1A	3.32 (1.77)	3.42 (1.50)	.568	b = .01, 95% CI: [-.03, .06]	b = .07, 95% CI: [-.02, .19]
Study 1B	3.28 (1.72)	4.12 (1.65)	< .001	b = .18, 95% CI: **[.06, .32]**	b = .31, 95% CI: **[.12, .51]**
Study 1C	3.81 (1.64)	4.33 (1.52)	= .001	b = -.04, 95% CI: [-.16, .06]	b = .26, 95% CI: **[.11, .47]**
Study 2	3.03 (1.40)	3.68 (1.46)	< .001	b = -.05, 95% CI: [-.17, .07]	b = .41, 95% CI: **[.23, .64]**
Study 3	3.52 (1.67)	3.56 (1.60)	.797	b = -.01, 95% CI: [-.03, .03]	b = .11, 95% CI: **[.01, .24]**
Study 4	4.07 (1.82)	3.67 (1.79)	< .001	b = .06, 95% CI: **[.01, .14]**	b = .14, 95% CI: **[.02, .27]**
Study 5*	3.17 (1.50)	3.60 (1.40)	.003	b = .02, 95% CI: [-.05, .11]	b = .34, 95% CI: **[.18, .53]**

*These results come only from the creative condition (rather than the combined fact and fiction conditions). In the fact condition specifically, the difference in quality between genAI and humans is non-significant (M_{human} = 4.67, SD = 1.42 vs. M_{genAI} = 4.87, SD = 1.28; F(1,399) = 2.21, p = .138).