ORIGINAL PAPER

Humans as Creativity Gatekeepers: Are We Biased Against AI Creativity?

Federico Magni[1] · Jiyoung Park[2] · Melody Manchi Chao[3]

Abstract

With artificial intelligence (AI) increasingly involved in the creation of organizational and commercial artifacts, human evaluators' role as creativity gatekeepers of AI-produced artifacts will become critical for innovation processes. However, when humans evaluate creativity, their judgment is clouded by biases triggered by the characteristics of the creator. Drawing from folk psychology and algorithm aversion research, we examine whether the identity of the producer of a given artifact as artificial intelligence (AI) or human is a source of bias affecting people's creativity evaluation of such artifact and what drives this effect. With four experimental studies ($N = 2039$), of which two were pre-registered, using different experimental designs and evaluation targets, we found that people sometimes—but not always—ascribe lower creativity to a product when they are told that the producer is an AI rather than a human. In addition, we found that people consistently perceive generative AI to exert less effort than humans in the creation of a given artifact, which drives the lower creativity ratings ascribed to generative AI producers. We discuss the implication of these findings for organizational creativity and innovation in the context of human-AI interaction.

Keywords Artificial intelligence · Creativity evaluation · Effort · Creativity · Cognitive biases · Human-AI interaction

Introduction

News such as the first auctioning of an AI-made artwork at Christie's in 2018, with the piece selling for a staggering US\$ 432,500 (Cohn, 2018), or the showcasing of the first AI-designed and 3D-printed chair at the famed Milan Design Week (Paciotti & Di Stefano, 2021) suggest that the creativity of AI-made artifacts is highly appreciated. However, the fear and anxiety of being replaced by generative AI (Schmelzer, 2019) and AI's limitations in terms of autonomy, motivation, and emotions (H.-K. Lee, 2022) often engender a negative assessment of AI-made creative artifacts (Castelo et al., 2019; Longoni et al., 2022; Prahl & Van Swol, 2017). So it is unclear if and how human evaluators rate the creativity of AI-generated ideas and products differently than that of comparable human-made artifacts. In other words, are humans good creativity gatekeepers?

Assessing creative production unbiasedly is challenging because creativity is often domain-specific, person-specific, and even situation-specific (Boden, 1998; Runco & Smith, 1992). Because of this, when assessing creativity, people often rely on heuristics and folk theories to guide their judgment (Baas et al., 2015; Ritter & Rietzschel, 2017) and are thus subject to evaluation biases (Beaty et al., 2018; Licuanan et al., 2007; Mastria et al., 2019). So far, we know about evaluators' biases in the assessment of creative artifacts made by humans—especially biases related to the identity of the producer (Baas et al., 2016; Rietzschel et al., 2016; Simonton, 2004). Basing one's creativity evaluations on such heuristics is problematic because

✉ Federico Magni
  fmagni@ethz.ch

  Jiyoung Park
  jiyoungpark@duksung.ac.kr

  Melody Manchi Chao
  mchao@ust.hk

[1] Department of Management, Technology and Economics, ETH Zürich, Weinbergstrasse 56-58, Zurich, Switzerland

[2] Duksung Women's University, 33 Samyang-Ro 144-Gil, Dobong-Gu, Seoul, South Korea

[3] Department of Management, Hong Kong University of Science & Technology, Clear Water Bay, Hong Kong SAR, China

Springer

the creativity of an artifact should be assessed as an intrinsic characteristic of the artifact itself, independent of who made it (Amabile, 1982, 1983, 2020). Thus, extant research on creativity evaluation focuses on how heuristics and biases obfuscate evaluations of the creativity of a target. However, we do not know whether the identity of a producer as a human or non-human agent—a factor external to the characteristic of the artifact—can also be a source of bias in creativity evaluation. The current study addresses this issue.

AI is defined as technology that can gather and interpret information, answer questions based on it, and evaluate its decision-making based on specific goals (Glikson & Woolley, 2020). With the booming of generative AI (Gozalo-Brizuela & Garrido-Merchan, 2023), the appraisal of AI creativity has become a pressing issue (Amabile, 2020). AI is now able to perform inherently creative tasks—from painting (Colton, 2012), writing poetry (Gervás, 2019), and composing music (du Sautoy, 2019) to formulating molecules for new drugs (Popova et al., 2018), designing architecture (Newton, 2019), and creating ideas for startups (https://ideasai.net/)—putting into question the monopoly of humans as idea generators.

As AI will increasingly perform generative tasks (Ferràs-Hernández, 2018), human involvement in the creative process will switch away from idea generation toward creativity evaluation, driving human workers to act as creativity gatekeepers. Humans will have the upper hand in creativity evaluation in the near future because, whereas AI excels at combining existing concepts and ideas—thereby generating novelty—it is still limited in the ability to understand when such novelty is also useful (and thus creative), especially in domains when such evaluation is complex and nuanced (Agrawal et al., 2017; Karimi et al., 2018).

Human-AI collaboration in creative production is already happening in specific organizational processes, such as the generation of technical advice for customers or the production of organizational artifacts, with AI generating solutions based on the combination of existing inputs and humans assessing the value and potential of these creations based on a holistic evaluation (van Esch et al., 2019; von Krogh, 2018). In the future of organizational creativity, AI will be increasingly involved in the generation of ideas, products, and solutions, and humans will increasingly act as creativity gatekeepers, evaluating the creativity of these artifacts. Thus, it is crucial to assess how apt humans are at evaluating AI creativity and specifically whether such evaluations are affected by biases related to the producer identity.

Tackling this issue is important because if one does not think that a certain artifact—be it a painting, a formula for a new drug, advice for solving a customer's problem, or a startup idea—is creative, one is less likely to support such artifact. Indeed, the evaluation of the creativity of ideas and products is a critical step of organizational innovation, as it impacts the implementation of creative ideas by influencing decision-making, resource assignment, and support (Amabile & Mueller, 2008; Loewenstein & Mueller, 2016). When an idea or a product does not receive strong support, it is unlikely to gather the necessary resources to achieve the implementation stage—e.g., founding a startup, developing a new product, selling a painting, or advice (Loewenstein & Mueller, 2016; Perry-Smith & Mannucci, 2017). So it is practically important to understand whether human evaluators discount the creativity of AI-made artifacts and, consequently, impede these artifacts from reaching the implementation stage and become actual innovation. This is consequential for organizations because innovation is key for survival and thriving in the modern business world (Anderson et al., 2014), and it would be costly to miss out on potential innovations because of discounting of the building blocks of innovations due to the identity of the producer as AI.

We not only hypothesize whether human evaluators rate artifacts described as AI-made (vs. human-made) as less creative, but also study how this might happen. Specifically, as people rate a given artifact production as requiring less effort when it is realized by AI rather than by humans (Bechwati & Xia, 2003; Chamberlain et al., 2018), we build on and expand the effort heuristic (Kruger et al., 2004) and posit that people's perception of lower effort exerted by AI is a mechanism driving the effect of producer identity on creativity evaluation. To test our hypotheses, we run four experimental studies (of which two were pre-registered; cumulative $N = 2039$) using a combination of within- and between-subjects design. The results of these studies identify whether human evaluators are appropriate gatekeepers of AI creativity in organizations and whether the effort heuristic biases their creativity evaluations when artifacts are produced by AI.

Theoretical Background and Hypotheses

Creativity Evaluation

Creativity research in organizational and social psychology has historically been defined as the production of novel and useful outputs and has widely focused on the idea-generation component of the creative process, to the point that creativity is often equated to and measured with idea generation (Amabile, 1983; Harvey & Berry, 2022; Paulus & Yang, 2000). Novelty and usefulness are defined in a broad sense, the former often used interchangeably with originality and unexpectedness and the latter with appropriateness, utility, quality, and effectiveness. Furthermore, this bi-dimensional conceptualization of creativity is sometimes integrated with additional dimensions, such as authenticity and aesthetics (Kharkhurin, 2014; Runco, 2004; Runco & Jaeger, 2012). Creativity dimensions—particularly novelty

and usefulness—can be positively, negatively, or not significantly related to each other (Harvey & Berry, 2022). Considering that we focus on evaluating the creativity of concrete artifacts and that the more "ideas are represented in more concrete form, the more an instance of creativity will fit the form of integration" (Harvey & Berry, 2022, p. 28), we adopt a creativity-as-integration perspective and expect the dimensions of creativity to be positively related to each other in our evaluators' eyes. This perspective, presupposing the shortest psychological distance between the creator/evaluator and the creation/evaluation context, is consistent with our research design, because participants directly interact with the artifacts and immediately evaluate their creativity, rather than engaging with distant and abstract alternative realities (see Harvey & Berry, 2022).

Moving beyond pure idea generation, researchers have increasingly acknowledged that the evaluation of creativity is crucial, as the acknowledgment of an idea's creativity by relevant evaluators is a necessary step toward its implementation and transformation into tangible innovation (Herman & Reiter-Palmon, 2011; Mueller et al., 2014; Perry-Smith & Mannucci, 2017; Ritter & Rietzschel, 2017). Creativity evaluators are thus gatekeepers of the creative process. However, humans are generally not very good at evaluating creativity (Mueller et al., 2012; Rietzschel et al., 2006, 2010).

The struggle in creativity evaluation is partly due to the nature of creativity. Creativity is often seen as a subjective concept (Boden, 1998; Loewenstein & Mueller, 2016; Luescher et al., 2019). It is difficult for people to objectively assess the creativity of an artifact, often because they lack clear criteria to guide their judgment about the intrinsic qualities of the artifact. Thus, they rely on assumptions and heuristics. Some of the factors impacting creativity evaluations are affective (Y. S. Lee et al., 2017; Mastria et al., 2019; Mueller et al., 2012), but most are imputable to cognitive biases that lead evaluators to infer creativity from factors such as personality traits and contextual cues (Baas et al., 2016; Boden, 2004; Ritter & Rietzschel, 2017).

The cognitive biases in the evaluation of creativity emerge because people use imprecise lay theories of creativity to guide their evaluations (Karwowski, 2014; Loewenstein & Mueller, 2016; Ritter & Rietzschel, 2017). Individual characteristics of the producer play a leading role in this regard. For instance, evaluators often resort to perceptions of the producer's age, mental and emotional stability, and "geniality" to guide creativity evaluations, even if these characteristics are often weakly linked to actual creativity (Baas et al., 2016; Boden, 2004; Ng & Feldman, 2008; Rietzschel et al., 2016; Simonton, 2004, 2014).

So far, research on producer-related biases in creativity evaluation has focused on human producers. However, thanks to steady technological advances, artificial agents have been increasingly able to perform creative production tasks that used to be considered the prerogative of humans (Colton, 2012; du Sautoy, 2019; Gervás, 2019; Oliveira, 2012; Popova et al., 2018). Surprisingly, little research has examined whether and how creativity evaluations are affected by the producer's identity as an artificial agent (Chamberlain et al., 2018; Moravčík et al., 2017). We aim to fill this research gap by conducting a theoretically grounded examination of the producer identity effect on creativity evaluation. Our arguments are grounded in folk psychology, as we examine how people's assessment of the actions and behaviors of artificial agents in general influence their evaluations of these agents' creativity.

Folk Psychology

Folk psychology studies the assessment and explanation of others' behaviors based on one's interpretation of others' minds, actions, and intentionality—the so-called folk theories (of mind, agency, intentionality, etc.) (Malle, 2011; Malle & Knobe, 1997). As described by the computers-are-social-actors framework (Nass & Moon, 2000; Reeves & Nass, 1996), people engage in *ethopoeia*—i.e., they apply folk theories to human–machine interactions, and they socially interact with artificial agents similarly as with other humans, as long as the artificial agents exhibit social cues that mirror human behavior (de Graaf & Malle, 2019; Thellman et al., 2017; von der Pütten et al., 2010). Consequently, we expect people to use folk theories to inform their creativity evaluations when they evaluate the creativity of artifacts produced by AI as they do when they evaluate artifacts produced by humans.

At the same time, human evaluators do perceive a difference of nature between human and artificial agents. The folk theory of artifact production informs us that "highlighting the role of humans in production processes can increase artifact value, relative to […] absence of people" (Judge et al., 2020, p. 7). Accordingly, artifacts and products that are (hand)made by humans are rated more favorably than comparable machine-made products (Abouab & Gomez, 2015; Fuchs et al., 2015), and people assign more value to a product when it is described as "made by people in a factory" than when it is simply described as "made in a factory" (Job et al., 2017).

We extend these theoretical and empirical arguments about the positive effect of human production on perceived value to the assessment of creative value specifically. We do so because people generally see artificial agents as unable to feel empathy and to read and manage emotions, as well as lacking a superordinate intentionality to guide their actions (Boden, 1998; Heer, 2019; Kim & Duhachek, 2020). As emotional management and superordinate intentionality are considered cornerstones of creativity, human evaluators' perception that artificial agents lack those skills would negatively affect their evaluation of the artificial agents' creative

ability and thus of the creativity of their productions (Boden, 1998; Hawley-Dolan & Winner, 2011; Hong, 2018).

Hypothesis 1: The producer's identity as AI (vs. human) has a negative effect on creativity evaluation.

The Effort Heuristic

Next, we investigate what might drive the effect of producer identity on creativity evaluation. The folk theory of artifact creation (Judge et al., 2020) points to effort as a key mechanism guiding people's evaluations of artifacts. To facilitate the evaluation of products, people tend to infer from their perception of the creation process, especially when objective and easily accessible criteria are lacking—which is often the case when evaluating creativity (Chinander & Schweitzer, 2003). Indeed, people assign value to perceived effort—the "physical labor, skill, and ingenuity" exerted by the producer in the creation process (Judge et al., 2020, p. 4)—and rate products more favorably when they perceive the creator to have invested a higher amount of effort in production (Buell & Norton, 2011; Mohr & Bitner, 1995; Newman & Bloom, 2012).

Relatedly, the effort heuristic (Kruger et al., 2004) informs the producer identity-creativity evaluation relationship because the perceptions of the effort exerted in a production process impact creativity evaluations and are at the same time influenced by producer identity. Indeed, human evaluators tend to perceive machines as generally exerting less effort than humans in performing tasks (Bechwati & Xia, 2003; Chamberlain et al., 2018; Kruger et al., 2004). Accordingly, the fact that something is *handmade* and produced by a human *artisan* activates the effort heuristic (Fuchs et al., 2015). Thus, taking together people's general assessment that AI exerts less effort than humans in a given production endeavor and extending the positive impact of higher effort perception on value appraisals to creative production, we expect the following:

Hypothesis 2: Perceived effort mediates the negative effect of producer identity as AI (vs. human) on creativity evaluation.

Overview of the Studies

We conducted four experimental studies to test our hypotheses. Our goal was to concurrently achieve internal consistency and generalizability. Therefore, we recruited samples from two regions (USA and Hong Kong) and from diverse populations (working adults and undergraduate students) and used different evaluation targets (visual and conceptual), different study designs (within- and between-person), and different creativity measures. All participants expressed their consent, were able to stop at any time, and were adequately debriefed.

On top of testing our hypotheses, we conducted additional analyses testing the effects of creativity evaluation on a behavioral outcome (willingness to pay) for studies 2–4. The results of these analyses are reported in Appendix 5.

Study 1

In study 1, we adopted a within-person design (producer identity: AI vs. human) to test hypothesis 1.

Methods

One hundred and thirty-one undergraduate business students from a Hong Kong university (66% female, $M_{age} = 19.05$, $SD_{age} = 1.26$) completed the survey in a lab in exchange for course credit. There were no missing data. Participants were told that they would be shown six Australian Aboriginal paintings and that "Some of these paintings have been created by aboriginal artists, based on their traditional painting style, while others have been created by artificial intelligence (AI) that has been trained to paint in a style consistent with the traditional aboriginal style." The paintings were divided into two groups: group 1 (paintings a, b, and c) and group 2 (paintings d, e, and f). Even though all paintings were comparable in terms of value and style (see Appendix 1 for more information), to account for any potential effect stemming from the content of the paintings, half of the participants were told that paintings in group 1 were produced by a human artist and paintings in group 2 by AI, and the other half of the participants were told the opposite. The order of presentation of the paintings was randomized. We asked participants to rate the creativity of each painting with a 3-item measure (*very uncreative* to *very creative*, *very inauthentic* to *very authentic*, and *very poor quality* to *very good quality*) to capture the multidimensional nature of creativity. We specifically decided to capture authenticity, a less commonly employed creativity subdimension, because the creative value of Australian Aboriginal paintings strongly relies on authenticity (Coleman, 2001). Inspired by past research (Cropley et al., 2011; Sullivan & Ford, 2010), we used *good quality* to assess usefulness because of the visual and approachable nature of our evaluation targets. All items here and in subsequent studies were measured on 5-point Likert scales. The scale exhibited good reliability (Cronbach's $\alpha = 0.81$), suggesting that the various subdimensions pointed to a consistent direction in line with the creativity-as-integration perspective (Harvey & Berry, 2022). To account for the multilevel nature of our data (painting evaluations

nested within individuals), we tested the hypothesis with a multilevel ANOVA using R nlme.

Results

We checked and found no significant difference in creativity evaluations between participants who were told that paintings in group 1 were painted by AI and those in group 2 by humans and participants who were told the opposite ($F(1, 655) = 3.09$, $p = 0.08$). Participants rated painting creativity lower when a painting was described as produced by AI ($M = 3.43$, $SD = 0.87$) rather than humans ($M = 3.65$, $SD = 0.80$), and the difference was significant ($F(1, 655) = 20.93$, $p < 0.001$), supporting hypothesis 1.

Study 2

Study 2 aimed to extend the finding of study 1 by testing perceived effort as a mechanism and by using a sample from a different population (American adults) and a different evaluation target (advertisement posters). Power analyses for this and subsequent studies are available in Appendix 4.

Methods

Participants

Participants of study 2 were USA-based individuals recruited on Amazon Mechanical Turk (MTurk) and were paid US$ 1 for completing the survey online, which took about 10 min. In the instructions for the study, we have included an attention check. Participants who failed the test were notified and not allowed to take part in the study. Besides that, we did not apply any specific filter to select participants. We requested a sample of 300 participants and obtained 303 participants (54% female, $M_{age} = 40.29$, $SD_{age} = 12.76$). Each participant rated five targets for a total of 1515 observations. There were no missing data, as participants were required to answer each question to the best of their abilities.

Procedure and Materials

In study 2, we adopted a between-person design: Participants were shown the manipulation, which assigned them to one of two conditions (producer identity: AI vs. human). In the AI (human) condition, they were told the following: "The posters have been designed by AI (a marketing designer) and are ready to be pitched to the supermarket's marketing director for final approval." To make the AI condition more believable, participants in that condition

were told that "The latest developments in machine learning have made it possible for AI to be trained in the design of original ads by being exposed to existing advertisement campaigns." Then, participants in both conditions were presented with identical evaluation targets. The advertisement posters were taken from the actual posters of an Italian supermarket's advertising campaign and are available upon request. The posters were cleaned from any written words, to make them equally accessible to all participants notwithstanding their spoken languages. Participants were shown the posters one by one and were asked to evaluate the creativity of each one. The order of presentation was randomized. After seeing all the posters, they were asked to rate their perception of the effort exerted by the producer in creating the posters and finally to report demographics and their familiarity with the domain of production and with AI.

Measures

Creativity Participants evaluated the posters' creativity with a 3-item measure (*very uncreative* to *very creative*, *very unoriginal* to *very original*, and *very poor quality* to *very good quality*), switching from *authentic* in study 1 to the more commonly used *original*, also because the evaluation targets (advertisement posters) were not inherently artistic. The scale showed good reliability (Cronbach's $\alpha = 0.86$), again supporting the creativity-as-integration meta-theory.

Perceived Effort We measured perceived effort with a scale composed of two items adapted from Bechwati and Xia (2003), "The artist (AI) put a lot of effort in the creation of the paintings" and "The artist (AI) worked hard in the creation of the paintings." The scale exhibited excellent reliability (Cronbach's $\alpha = 0.96$).

Controls In the analyses, we included as controls some variables that are commonly included in creativity evaluation studies, such as age, gender, and education, as well as familiarity with the domain of production, measured with two items ("How familiar are you with marketing campaigns?" and "How familiar were you with the design of advertisement posters before taking this survey?"), and familiarity with AI, measured with a single item "How familiar are you with AI?" on a 5-point familiarity Likert scale. We controlled for familiarity with the domain of production and with AI because participants who are highly familiar might be less biased because their deeper knowledge allows them to reduce the ambiguity in creativity evaluations (Park & Lessig, 1981) and also as individuals might have a preference for familiarity or for unfamiliarity in different contexts (Liao et al., 2011).

Analytical Strategy

As in study 1, our data had a nested structure (poster evaluations nested within individuals). Thus, we tested hypothesis 1 in the same way as in study 1. For hypothesis 2, we ran a multilevel regression in MPlus using the type = *complex* command to account for data nestedness and then estimated 5000 bootstrapping 95% confidence intervals (CI) in R to test the mediation hypothesis (Preacher & Selig, 2012). Correlations and regression results for study 2 are reported in Appendix 3. Significance patterns were consistent in testing the models with and without control variables.

Results

We did not find a significant difference ($F(1, 301) = 0.12$, $p = 0.726$) in the creativity evaluations of the posters between the AI producer ($M = 3.83$, SD = 1.00) and the human producer ($M = 3.86$, SD = 0.94) conditions, not supporting hypothesis 1. There was instead a significant indirect effect of producer identity on creativity evaluation through perceived effort (bootstrapped $b = 0.29$, CI: [0.210, 0.378]), supporting hypothesis 2, as AI ($M = 4.11$, SD = 1.63) was perceived to exert less effort than humans ($M = 5.32$, SD = 1.32) in the production of the posters, and perceived effort positively related to creativity ($b = 0.25$, $p < 0.001$). Hypothesis 2 was thus supported because mediation can exist even when the direct effect is non-significant (Rucker et al., 2011; Zhao et al., 2010).

Study 3

In study 3, which was pre-registered (see Appendix 2), we expand the previous findings by using a non-visual evaluation target (i.e., business ideas) and by delving into different dimensions of perceived effort.

Methods

We requested a sample of 800 participants on MTurk and obtained 802 participants (57% female, $M_{age} = 43.85$, $SD_{age} = 13.76$). The experimental design and participant recruitment conditions were the same as in study 2, and participants from study 2 were not allowed to participate in the study to make sure that each participant would start the survey naïve about the manipulation. Participants in the AI (human) condition were told that they would "read a series of ideas for new startup businesses that are ready to be evaluated by a venture capitalist. These ideas have been generated by a text-processing AI (an aspiring entrepreneur), who was exposed to countless startup ideas before coming up with these." They were also provided with either of the following definitions of AI and entrepreneurs in line with the randomly assigned experimental condition: "AI is defined as technology that can gather and interpret information, produce outputs, and evaluate its decision-making based of specific goals, similarly to the way a human mind does" or "Entrepreneurs are defined as individuals who set up businesses taking on risks in the hope of generating profit."

The business ideas that participants were asked to evaluate were actually created by AI: We took five ideas from the homepage of https://ideasai.net/, a website that uses the AI deep learning model GPT-3 to produce ideas for business startups. A sample idea was "A platform to help people find the best care for the elderly, with a focus on home healthcare."

Creativity evaluations were measured following the same theoretical rationale as in study 2 in conceptualizing creativity as an integrative construct. The 3-item measure and scales (*not at all creative* to *very creative*, *not at all novel* to *very novel*, *not at all appropriate* to *very appropriate*; Cronbach's $\alpha = 0.76$) was inspired from extant research (De Dreu, 2010; Heinen & Johnson, 2018) and adapted from studies 1 and 2 to increase the generalizability of findings and to cater the items to the evaluation target. Specifically, we switched from *good quality* to *appropriate* as a measure of the usefulness dimension because the latter is commonly used in the evaluation of ideas (e.g., Kleinmintz et al., 2019).

To dig deeper into how participants envisioned effort when evaluating their effort perceptions of the idea-generation processes, we extended the previous two-item scale with additional elements tackling important effort dimensions identified by past research (Kruger et al., 2004; Massin, 2017): temporal effort ("[…] took a long time to come up with the ideas"), physical effort ("[…] exerted substantial physical effort to come up with the ideas"), cognitive effort ("[…] invested a lot of cognitive resources in coming up with the ideas"), and motivational effort ("[…] was strongly motivated to come up with the ideas"). The six-item scale including the previous two items and the new four items showed excellent reliability (Cronbach's $\alpha = 0.93$), and an exploratory factor analysis indicated that all items load on a single factor (with all loadings higher than 0.74), pointing to a unique latent effort construct.

Control variables were the same as in study 2, except for adapting the two items measuring domain familiarity to the context of business ideas ("How familiar are you with entrepreneurship?", "How familiar were you with the context of venture capital and startups before taking this survey?"). The analytical strategy was the same as in study 2. Correlations and regression results for study 3 are reported in Appendix 3.

Results

We found no support for hypothesis 1: Participants' creativity evaluations were not significantly affected by the producer identity ($F(1, 800) = 0.08$, $p = 0.783$), as there was no significant difference between the human condition ($M = 3.43$, SD = 0.90) and the AI condition ($M = 3.45$, SD = 0.93). As in study 2, we instead found evidence supporting hypothesis 2, as effort perceptions mediated the effect of producer identity on creativity evaluation (bootstrapped $b = 0.25$, CI: [0.202, 0.310]). Specifically, AI ($M = 3.38$, SD = 1.36) was perceived as exerting less effort than humans ($M = 4.56$, SD = 1.34) in the production of business ideas and perceived effort positively related to creativity ($b = 0.22$, $p < 0.001$).

Additional Analyses

We conducted additional analyses using alternative measures of effort: the two-item scale used in study 2 and each additional single-item as a standalone measure. Results were consistent with any effort measure: ANOVA tests showed that participants rated humans to exert more effort with any measure at the $p < 0.001$ significance level, and all measures mediated the effect of producer identity on product creativity evaluation (2-item effort: bootstrapped $b = 0.29$, CI: [0.234, 0.343]; temporal effort: bootstrapped $b = 0.21$, CI: [0.159, 0.252]; physical effort: bootstrapped $b = 0.14$, CI: [0.101, 0.179]; cognitive effort: bootstrapped $b = 0.18$, CI: [0.138, 0.230]; motivational effort: bootstrapped $b = 0.24$, CI: [0.194, 0.294]). We thus inferred that, notwithstanding what specific dimension of effort participants have in mind when completing the experiment, they consistently rate humans as exerting more effort than AI in the performance of a given task.

Study 4

In study 4, which was pre-registered (see Appendix 2), we complement the previous findings by manipulating not only producer identity but also the effort exerted in the production process, so as to test the causal relation between effort perception and creativity evaluation (Pirlott & MacKinnon, 2016).

Pilot Study

We manipulated effort using time spent by the creator to complete an artifact, following Kruger and colleagues (2004). To ensure that the time-based manipulation would produce the intended effects (i.e., influence evaluators' effort perception), we first pilot-tested it. We recruited 100 participants on MTurk, showed them each of the five paintings according to a condition in a 2 (producer identity: AI vs. human) × 2 (effort: low vs. high) design, and asked them to report their perceptions of the effort exerted by the producer with the item "This painting took a lot of effort to complete," as well as the painting's creativity with the item "This painting is very creative." The manipulation worked, as we found that participants rated the paintings as requiring more effort ($F(1, 399) = 66.71$, $p < 0.001$) when assigned to the high-effort condition ($M = 6.25$, SD = 0.89) than when assigned to the low-effort condition ($M = 5.65$, SD = 1.26). We also got initial support for the expectation that effort causally impacted creativity, as paintings in the high-effort condition were rated on average as more creative ($M = 6.05$, SD = 1.14) than paintings in the low-effort condition ($M = 5.86$, SD = 1.02; $F(1, 399) = 4.88$, $p < 0.05$).

Methods

After validating the effort manipulation, we requested a sample of 800 participants on MTurk and obtained 808 participants (54% female, $M_{age} = 40.25$, $SD_{age} = 11.94$). The participants were recruited and attention check filtered under the same conditions as in previous studies, and participants from previous studies were not allowed to participate in the study. The experimental design was similar to the previous studies but with two main differences: The design was a 2 (producer identity: AI vs. human) × 2 (effort: low vs. high) factorial, and the assignment of each condition was made at a painting (rather than individual) level, resulting in a within-person design. We used the same effort single-item measure as in the pilot as a manipulation check. The paintings shown to participants were five of the six used in study 1 (see Appendix 1 for details). Participants were told that "The latest developments in AI have made it possible for AI to paint according to the Aboriginal artists' style after seeing a number of their paintings. AI is defined as technology that can gather and interpret information, produce outputs, and evaluate its decision-making based of specific goals, similarly to the way a human mind does." Participants evaluated the paintings' creativity with the same scale as in study 3 (Cronbach's $\alpha = 0.83$). Control variables were the same as in study 2, except for adapting the two items measuring domain familiarity to the context of Australian Aboriginal paintings ("How familiar are you with figurative art?", "How familiar were you with Australian Aboriginal art before taking this survey?").

Given the manipulation-of-mediator design, the analytical strategy for study 4 did not involve mediation analyses. Thus, we checked multilevel multivariate (producer identity and effort) ANOVA, and we tested the effects on creativity evaluation of producer identity and of effort individually. We further tested on MPlus their individual and interactive effects in a multilevel regression including the control variables. Correlations and regression results for study 4 are reported in Appendix 3.

Results and *Discussion*

As in the pilot study, our manipulation worked: Paintings in the high-effort conditions ($M = 5.52$, SD $= 1.53$) were rated as requiring higher effort ($F(1, 3,231) = 232.44$, $p < 0.001$) than paintings in the low-effort conditions ($M = 4.86$, SD $= 1.65$). We found support for hypothesis 1, as participants rated the paintings as more creative ($F(1, 3,230) = 447.28$, $p < 0.001$) when they were presented as produced by a human ($M = 4.27$, SD $= 0.70$) rather than by an AI ($M = 3.84$, SD $= 0.88$). At the same time, the effort manipulation also had an effect on creativity evaluation as hypothesized: Paintings in the high effort conditions ($M = 4.11$, SD $= 0.81$) were rated as more creative ($F(1, 3,230) = 37.54$, $p < 0.001$) than those in the low-effort conditions ($M = 4.01$, SD $= 0.83$). This finding showed the causal effect of effort on creativity in line with hypothesis 2. Effect sizes on effort and creativity for all studies are reported in Table 1; further information on their calculation is reported in Appendix 4.

We further conducted a post hoc Tukey test on SPSS to check whether the creativity ratings of pairs of conditions were significantly different from each other. This analysis showed that all pairs were significantly different from each other, forming a ranking of conditions: Human-high effort led to the highest creativity ratings, followed by human-low effort (mean difference $= 0.11$, $p < 0.05$), in turn followed by AI-high effort (mean difference $= 0.33$, $p < 0.001$), and in turn followed by AI-low effort (mean difference $= 0.10$, $p < 0.05$). Of particular interest is the fact that the human-low-effort condition led to higher creativity ratings than the AI-high-effort condition. This means that effort does not substitute the effect of the producer identity (which would be shown by a not significant difference between the human-high-effort and AI-high-effort conditions, as well as between the human-low-effort and AI-low-effort conditions), but

Table 1 Effect sizes

Variables	Perceived effort (*d*)	Creativity evaluation (*d*)	Mediation (f^2)
Study 1 ($n = 786$)			
Producer identity	—	0.261	—
Study 2 ($n = 1515$)			
Producer identity	0.818	0.028	0.164
Study 3 ($n = 4010$)			
Producer identity	0.868	-0.012	0.116
Study 4 ($n = 4040$)			
Producer identity	—	0.541	—
Effort	—	0.125	—

$n =$ sample size at observation level; $d =$ Cohen's *d* for producer identity (AI $= 0$, human $= 1$) on outcomes; $f^2 =$ Cohen's f^2 for the mediation of perceived effort on the effect of producer identity on creativity evaluation

rather the effect of identity prevails on that of effort, suggesting that the mediating role of effort is partial.

The multilevel regression analyses showed consistent results with the ANOVAs, as both the producer identity manipulation ($b = 0.43$, $p < 0.001$) and the effort manipulation ($b = 0.10$, $p < 0.001$) impacted the paintings' creativity evaluation as hypothesized. However, the two manipulations did not interactively predict creativity ($b = 0.01$, $p = 0.926$). Consistent with the Tukey test, this indicates that the effect of producer identity on creativity was not different at high effort or low effort, suggesting that effort does not completely explain the producer identity effect on creativity and thus acts as a partial (rather than full) mediator.

General Discussion

The main objective of the present research was to understand whether human evaluators are good gatekeepers of AI creativity. We tested the effect of producer identity as AI (vs. human) on creativity evaluation and the mediating role of perceived effort in this relationship. Taking together the results of four studies on a cumulative sample of 2039 participants, we found that people *only sometimes* assign lower creativity to a given production when they are told that AI created it compared to when they are told that a human created it (Chamberlain et al., 2018; Hong, 2018; Kirk et al., 2009; Moffat & Kelly, 2006). Thus, it appears that human evaluators are better creativity gatekeepers in some circumstances than in others, potentially depending on factors such as the nature of the evaluation target. Specifically, the existence of a direct producer identity bias was evidenced in two studies that used artistic visual evaluation targets (Australian Aboriginal paintings), but not in two other studies that asked people to evaluate advertisement posters and ideas for business startups. These findings not only contribute to the nascent research of human evaluation of AI creativity, but also extend folk psychology, in particular the folk theory of artifact creation (Judge et al., 2020) and the effort heuristic (Kruger et al., 2004), by applying them to the realm of creativity evaluation.

Effort as a Key Mechanism

We found consistent evidence that effort is a key mechanism driving the effect of producer identity on creativity evaluation: All three studies in which we tested effort—with different measures, designs, and evaluation targets—clearly indicated that human evaluators appraise a given production process as less effortful when it is carried out by AI rather than by humans, and this leads them to assess the creativity of the resulting product as lower, supporting the expansion of the effort heuristic to the domain of creative production. This

finding on the one hand corroborates the relevance of effort perceptions as a key mechanism and on the other hand raises further questions about why people perceive artificial agents to exert less effort than humans in the creation of a given object. Could it be that evaluators discount the effort expended by the artificial agent's creator and see the machine as an efficient tool capable of effortless production? Or is it because artificial agents are perceived as unable to feel and control emotions and to be guided by a superordinate intentionality? In other words, are these typically humane characteristics the drivers of the differential effort perceptions? Investigating the main dimensions of effort, such as time invested, physical effort, cognitive effort, and motivational effort (Kruger et al., 2004; Massin, 2017), we found that these dimensions shared a large amount of variability and were all equally able to capture the mediating effect of effort on the link between producer identity and creativity evaluation. More research is needed to disentangle this knot.

Another interesting question to address in relation to effort perceptions is about the evaluator's exposure to the creation process. Past studies showed that when people see with their own eyes how artificial agents create an artifact, their responses are more favorable, and they perceive higher effort expended in production and higher product value in general (Buell et al., 2017; Chamberlain et al., 2018). It would thus be interesting, moving forward, to assess whether exposure to the creation process and collaboration with AI impacts the effect of producer identity on creativity evaluations (Colton, 2008), especially in light of the rising diffusion of generative AI (Brynjolfsson et al., 2023; Noy & Zhang, 2023) and the ensuing sense of reciprocity and trust toward AI (Buell & Norton, 2011; Glikson & Woolley, 2020).

Another point of reflection about effort relates to the ongoing (if somewhat implicit) debate in the literature about whether people evaluate as more creative producers (and ensuing products) who put a lot of effort in their creation process or those who instead put very little effort and are perceived to achieve great outcomes mostly thanks to their innate abilities. In our research, consistent with folk theories of artifact creation (Judge et al., 2020), we found both correlational and causal support for the former argument in the context of evaluating artifact production. Effort seems to positively inform creativity evaluations in the context of (AI vs. human) artifact creation. However, elsewhere, evidence has been found in support of the opposite argument—i.e., that perceptions of higher effort invested in the production process lead to lower creativity ratings (Tsay, 2016). We do not exclude that in other circumstances and domains of production, this negative effect might exist. The question is not only an empirical but also a theoretical and philosophical one: As researchers of creativity in organizations have been striving for decades to reframe the definition of creativity from innate talent and genius to something that everyone

can have and develop to some extent (Amabile, 2006), the question of the link between effort perceptions and creativity cannot be separated from the very definition of creativity. Is creativity the result of talent, or is it the result of effort? And, most interestingly, how does this debate apply to the creative production by AI? Likely, both components—talent and effort—play a part, and either one can dominate in any given context. Yet both effort and talent are fuzzy concepts when applied to artificial agents, and we do not know much about how people assess AI's talent and effort. We encourage future research to inquire on this fundamental issue.

Characteristics of AI

The research design and manipulations we used were relatively simple. This is a point of strength when testing a baseline, generic effect, as in this case. Yet a limitation of such a design is that it is unable to account for several nuances. While our research employed a generic form of generative AI that creates texts and images, we did not focus on how specific features of AI as a producer might influence creativity evaluations of its production. Recent research has suggested that different AI representations—e.g., robotic AI with a physical presence, virtual AI perceived by end-users as a virtual agent, and embedded AI largely invisible to end-users—have different impacts on people's assessments of its behaviors (Glikson & Woolley, 2020; Malle et al., 2019). Future research might study whether AI representation is a boundary condition of the producer identity bias in creativity evaluation.

Another AI characteristic we did not consider in our design is the extent of AI's anthropomorphism. Whether artificial agents assume a humanoid appearance is likely to impact people's reactions to them and to their productions (Chamberlain et al., 2018; Glikson & Woolley, 2020; Waytz et al., 2014). This is particularly relevant when considering that people tend to relate to artificial agents in a very similar way as to other humans, as suggested by folk psychology (de Graaf & Malle, 2017; Thellman et al., 2017). It is thus likely that the more an intelligent artificial agent assumes a human-like form, the more closely people will treat it to an actual human agent. Thus, anthropomorphism might moderate the extent of the producer identity effect on judgments and assessments, including creativity evaluation. Depending on eventual findings, this direction for future research also offers possibilities for designing interventions aimed at reducing the producer identity bias.

Other Domains of Production

We saw that the target of production seems to influence the effect of producer identity on product creativity evaluation. In particular, we found that people discount AI creativity

for more aesthetic production (e.g., Australian Aboriginal paintings), but not for more commercial production (e.g., advertisement posters, startup ideas). From this, we tentatively suggest that the domain of production may be a switch turning on (in aesthetic production) or off (in commercial production) the impact on creativity evaluations of the identity of the producer as human or AI. It would be interesting to further test this relationship in various domains of creative production. For instance, AI is currently being used to perform creative tasks in the scientific domain—e.g., to develop new compounds for medicines (Fleming, 2018; Popova et al., 2018). How do people evaluate AI creativity in this domain? Is creativity seen as a re-combination of existing components and the result of a trial-and-error process, or does the acknowledgment of creativity stem from the producer's ability to see a holistic picture and to challenge existing criteria of evaluation? In the former case, AI production might be considered as (if not more) creative than humans' due to AI's superior computational ability; in the latter case, AI's perceived inability to go beyond a given set of instructions might instead lead its work to be deemed as less creative.

Practical Implications

Our research findings suggest that it is advisable to consider intervention programs (e.g., see Burton et al., 2020; Longoni & Cian, 2022) to raise awareness about human biases toward AI-generated creative outcomes and make human evaluators better creative gatekeepers. Targeted interventions should address when (e.g., for productions framed as more aesthetic) and why (e.g., because of lower effort perceptions) biases against generative AI occur as well as raise awareness among evaluators so as to prevent them.

Another important consideration in adopting generative AI relates to workers' willingness to disclose the use of AI due to concerns about creativity evaluation. While people may fear that their creations may be discriminated if they disclose a collaboration with generative AI, new domains of human-AI creativity may emerge, such as prompt engineering—i.e., the search for prompts that allow generative AI to produce desired outputs (Liu & Chilton, 2022). Creating awareness of potential changes and new forms of creativity arising from AI adoption, organizations need to establish ethical and practical guidelines on how to utilize and evaluate AI to ensure transparency and fairness.

Additionally, our findings have direct implications for marketing decision-making. The integration of generative AI into the creative process can lead customers to undervalue the efforts invested in the resulting outcomes. When marketing and communicating the creative outcomes of AI-human collaboration, marketers can reduce the bias against AI creativity by emphasizing human involvement

in the process, human-like features of AI, and the effort dedicated to the creative process (Castelo et al., 2019; Chamberlain et al., 2018).

Limitations

Limitations of this research, in addition to what has been discussed above, include the following. First, due to the paucity of empirical research on creativity evaluation as well as to the complex and multidimensional nature of creativity, we derived and inspired our measures of artifact creativity from extant theories and empirical studies (e.g., Harvey & Berry, 2022; Runco, 2004; Sullivan & Ford, 2010). Specifically, we adopted—and found empirical support for—the creativity-as-integration approach to assess creativity as an integrative construct with multiple dimensions positively related to each other (Harvey & Berry, 2022). This approach is appropriate because (1) it is consistent with the dominant theoretical framework and empirical measures in the creativity literature and (2) it is consistent with how evaluation targets (e.g., consumer products, employee outputs) are commonly assessed (e.g., George & Zhou, 2001; Oldham & Cummings, 1996; Zhou et al., 2019).

A second possible limitation is that participants were told that AI was involved in the production process in general, but they were given no specific information about the exact extent of AI's involvement in the production process. The instructions we gave them focused on the involvement of AI in artifact production in a general sense because we examined an emergent phenomenon. The primary goal of our research was to test the general impact of having an AI producer on creativity evaluation, rather than to narrowly focus on a specific facet of the phenomenon. Nonetheless, we encourage future research to explore how different extents of AI involvement in creative processes might have different effects on how AI creativity is evaluated.

Another potential limitation is that the participants may have different ideas about the characteristics and abilities of AI. Given that AI capabilities have been developing quickly thanks to fast technological innovation cycles (Vinuesa et al., 2020), it is important to clarify to participants what one means with AI. In study 1, we did not elaborate on a definition of AI and rather left this open for participants, which might be a limitation as participants might have different ideas about what an AI is and what it can do. To remedy this, in subsequent studies, we have more clearly provided participants with definitions and descriptions of AI. We expect that most lay people would have a generally shared and superficial understanding about AI and what it can do; nonetheless, in future studies, it would be important to explicitly describe the abilities of the AI under analysis, especially because technology has been advancing rapidly and the variance among laypeople conceptualizations about AI may increase.

Conclusions

With four experimental studies, we tested whether people evaluate differently the creativity of productions when these productions are described as produced by AI or by humans. Human evaluators only sometimes judged a product to be less creative when they were told that AI created it rather than a human. Specifically, the effect manifested itself when evaluating certain kinds of products, such as paintings, but not others, such as advertisement posters and business ideas. Consistent evidence instead was found about perceptions of exerted effort mediating this effect, as AI was perceived to exert less effort than humans in the production of a given output, which led to perceptions of lower creativity.

Our results substantiate previous findings on the producer identity bias only in part, as they show that the bias is not ubiquitous and that human evaluators' inadequacy as creativity gatekeepers seems to be situated in the context of AI creative production for aesthetic purposes. We also theoretically extend the extant literature by applying a folk psychology framework to the sphere of AI creative production. This work takes a pioneering step in analyzing people's evaluation of AI creativity and offers a springboard for future research to further study how the identity of the producer as an artificial (rather than human) agent impacts people's evaluations of creative productions.

Declarations

References

Abouab, N., & Gomez, P. (2015). Human contact imagined during the production process increases food naturalness perceptions. *Appetite, 91*, 273–277. https://doi.org/10.1016/j.appet.2015.04.002

Agrawal, A., Gans, J., & Goldfarb, A. (2017). *What to expect from artificial intelligence*. MIT Sloan Management Review.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.43.5.997

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology, 45*(2), 357–376. https://doi.org/10.1037/0022-3514.45.2.357

Amabile, T. M. (2006). How to kill creativity. *Creative Management and Development, Third Edition*. https://doi.org/10.4135/9781446213704.n2

Amabile, T. M. (2020). Creativity, artificial intelligence, and a world of surprises. *Academy of Management Discoveries, 6*(3), 351–354. https://doi.org/10.5465/amd.2019.0075

Amabile, T. M., & Mueller, J. S. (2008). Studying creativity, its processes, and its antecedents: An exploration of the componential theory of creativity. In *Handbook of Organizational Creativity*.

Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A State-of-the-science review, prospective commentary, and guiding framework. *Journal of Management, 40*(5), 1297–1333. https://doi.org/10.1177/0149206314527128

Baas, M., Koch, S., Nijstad, B. A., & De Dreu, C. K. W. (2015). Conceiving creativity: The nature and consequences of laypeople's beliefs about the realization of creativity. *Psychology of Aesthetics, Creativity, and the Arts, 9*(3), 340–354. https://doi.org/10.1037/a0039420

Baas, M., Nijstad, B. A., Boot, N. C., & De Dreu, C. K. W. (2016). Mad genius revisited: Vulnerability to psychopathology, biobehavioral approach-avoidance, and creativity. *Psychological Bulletin, 142*(6), 668–692. https://doi.org/10.1037/bul0000049

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., Fink, A., Qiu, J., Kwapil, T. R., Kane, M. J., & Silvia, P. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings of the National Academy of Sciences of the United States of America, 115*(5), 1087–1092. https://doi.org/10.1073/pnas.1713532115

Bechwati, N. N., & Xia, L. (2003). Do computers sweat? The impact of perceived effort of online decision aids on consumers' satisfaction with the decision process. *Journal of Consumer Psychology, 13*(1–2), 139–148. https://doi.org/10.1207/153276603768344852

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence, 103*(1–2), 347–356. https://doi.org/10.1093/acprof:oso/9780199836963.003.0012

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Psychology Press.

Brynjolfsson, E., Buffington, C., Goldschlag, N., Li, J. F., Miranda, J., & Seamans, R. (2023). Robot hubs: The skewed distribution of robots in US manufacturing. *AEA Papers and Proceedings, 113*, 215–218.

Buell, R. W., & Norton, M. I. (2011). The labor illusion: How operational transparency increases perceived value. *Management Science*. https://doi.org/10.1287/mnsc.1110.1376

Buell, R. W., Kim, T., & Tsay, C. J. (2017). Creating reciprocal value through operational transparency. *Management Science, 63*(6), 1673–1695. https://doi.org/10.1287/mnsc.2015.2411

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220–239. https://doi.org/10.1002/bdm.2155

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

Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts, 12*(2), 177–192. https://doi.org/10.1037/aca0000136

Chinander, K. R., & Schweitzer, M. E. (2003). The input bias: The misuse of input information in judgements of outcomes. *Organizational Behavior and Human Decision Processes, 91*(2), 243–253. https://doi.org/10.1016/S0749-5978(03)00025-6

Cohn, G. (2018). AI art at Christie's sells for $432,500. *The New York Times*. https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html. Accessed 29 Jun 2022.

Coleman, E. B. (2001). Aboriginal painting: Identity and authenticity. *The Journal of Aesthetics and Art Criticism, 59*(4), 385–402.

Colton, S. (2008). Creativity versus the perception of creativity in computational systems. *AAAI Spring Symposium - Technical Report, SS-08–03*, 14–20.

Colton, S. (2012). The painting fool: Stories from building an automated painter. In *Computers and Creativity* (pp. 3–38). Springer. https://doi.org/10.1007/978-3-642-31727-9_1

Cropley, D. H., Kaufman, J. C., & Cropley, A. J. (2011). Measuring creativity for innovation management. *Journal of Technology Management & Innovation, 6*(3), 13–30.

De Dreu, C. K. (2010). Human creativity: Reflections on the role of culture. *Management and Organization Review, 6*(3), 437–446.

du Sautoy, M. (2019). Can AI ever be truly creative? *New Scientist, 242*(3229), 38–41. https://doi.org/10.1016/S0262-4079(19)30840-1

Ferràs-Hernández, X. (2018). The future of management in a world of electronic brains. *Journal of Management Inquiry, 27*(2), 260–263. https://doi.org/10.1177/1056492617724973

Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature, 557*(7706), S55+. https://doi.org/10.1038/d41586-018-05267-x

Fuchs, C., Schreier, M., & van Osselaer, S. M. J. (2015). The handmade effect: What's love got to do with it? *Journal of Marketing, 79*(2), 98–110. https://doi.org/10.1509/jm.14.0018

George, J. M., & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: An interactional approach. *Journal of Applied Psychology, 86*(3), 513–524. https://doi.org/10.1037/0021-9010.86.3.513

Gervás, P. (2019). Exploring quantitative evaluations of the creativity of automatic poets. In F. Veale, T., Cardoso (Ed.), *Computational Creativity. Computational synthesis and creative systems* (pp. 275–304). Springer, Cham. https://doi.org/10.1007/978-3-319-43610-4_13

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A state of the art review of large generative AI models. *ArXiv Preprint ArXiv:2301.04655*.

de Graaf, M. M. A., & Malle, B. F. (2017). How people explain action (and autonomous intelligent systems should too). *AAAI Fall Symposium - Technical Report, FS-17–01-*, 19–26.

de Graaf, M. M. A., & Malle, B. (2019). People's explanations of robot behavior subtly reveal mental state inferences. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 239–248.

Harvey, S., & Berry, J. W. (2022). *Toward a meta-theory of creativity forms: How novelty and usefulness shape creativity*.

Hawley-Dolan, A., & Winner, E. (2011). Seeing the mind behind the art: People can distinguish abstract expressionist paintings from highly similar paintings by children, chimps, monkeys, and elephants. *Psychological Science, 22*(4), 435–441. https://doi.org/10.1177/0956797611400915

Heer, J. (2019). Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences of the United States of America, 116*(6), 1844–1850. https://doi.org/10.1073/pnas.1807184115

Heinen, D. J., & Johnson, D. R. (2018). Semantic distance: An automated measure of creativity that is novel and appropriate. *Psychology of Aesthetics, Creativity, and the Arts, 12*(2), 144.

Herman, A., & Reiter-Palmon, R. (2011). The effect of regulatory focus on idea generation and idea evaluation. *Psychology of Aesthetics, Creativity, and the Arts, 5*(1), 13–20. https://doi.org/10.1037/a0018587

Hong, J.-W. (2018). Bias in perception of art produced by artificial intelligence. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 290–303. https://doi.org/10.1007/978-3-319-91244-8

Job, V., Nikitin, J., Zhang, S. X., Carr, P. B., & Walton, G. M. (2017). Social traces of generic humans increase the value of everyday objects. *Personality and Social Psychology Bulletin, 43*(6), 785–792. https://doi.org/10.1177/0146167217697694

Judge, M., Fernando, J. W., Paladino, A., & Kashima, Y. (2020). Folk theories of artifact creation: How intuitions about human labor influence the value of artifacts. *Personality and Social Psychology Review, 24*(3), 195–211. https://doi.org/10.1177/1088868320905763

Karimi, P., Grace, K., Maher, M. L., & Davis, N. (2018). *Evaluating creativity in computational co-creative systems*. http://arxiv.org/abs/1807.09886

Karwowski, M. (2014). Creative mindsets: Measurement, correlates, consequences. *Psychology of Aesthetics, Creativity, and the Arts, 8*(1), 62–70. https://doi.org/10.1037/a0034898

Kharkhurin, A. V. (2014). Creativity. 4in1: Four-criterion construct of creativity. *Creativity Research Journal, 26*(3), 338–352.

Kim, T. W., & Duhachek, A. (2020). Artificial intelligence and persuasion: A construal-level account. *Psychological Science, 31*(4), 363–380. https://doi.org/10.1177/0956797620904985

Kirk, U., Skov, M., Hulme, O., Christensen, M. S., & Zeki, S. (2009). Modulation of aesthetic value by semantic context: An fMRI study. *NeuroImage, 44*(3), 1125–1132. https://doi.org/10.1016/j.neuroimage.2008.10.009

Kleinmintz, O. M., Ivancovsky, T., & Shamay-Tsoory, S. G. (2019). The two-fold model of creativity: The neural underpinnings of the generation and evaluation of creative ideas. *Current Opinion in Behavioral Sciences, 27*, 131–138.

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology, 40*(1), 91–98. https://doi.org/10.1016/S0022-1031(03)00065-9

Lee, H.-K. (2022). Rethinking creativity: Creative industries, AI and everyday creativity. *Media, Culture & Society, 44*(3), 601–612.

Lee, Y. S., Chang, J. Y., & Choi, J. N. (2017). Why reject creative ideas? Fear as a driver of implicit bias against creativity. *Creativity Research Journal, 29*(3), 225–235. https://doi.org/10.1080/10400419.2017.1360061

Liao, H. I., Yeh, S. L., & Shimojo, S. (2011). Novelty vs. familiarity principles in preference decisions: Task-context of past

experience matters. *Frontiers in Psychology, 2*(43), 1–8. https://doi.org/10.3389/fpsyg.2011.00043

Licuanan, B. F., Dailey, L. R., & Mumford, M. D. (2007). Idea evaluation: Error in evaluating highly original ideas. *Journal of Creative Behavior, 41*(1), 1–27. https://doi.org/10.1002/j.2162-6057.2007.tb01279.x

Liu, V., & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–23.

Loewenstein, J., & Mueller, J. (2016). Implicit theories of creative ideas: How culture guides creativity assessments. *Academy of Management Discoveries, 2*(4), 320–348. https://doi.org/10.5465/amd.2014.0147

Longoni, C., & Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The "word-of-machine" effect. *Journal of Marketing, 86*(1), 91–108.

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). News from generative artificial intelligence is believed less. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 97–106.

Luescher, R., Barthelmess, P. Y. Z., Kim, S. Y., Richter, U. H., & Mittag, M. (2019). Conceptualizing creativity: General and cultural biases in Gough's creative personality scale. *Journal of Creative Behavior, 53*(1), 30–43. https://doi.org/10.1002/jocb.160

Malle, B. F., Thapa Magar, S., & Scheutz, M. (2019). AI in the sky: How people morally evaluate human and machine decisions in a lethal strike dilemma. In *Robotics and Well-Being* (pp. 111–133). Springer, Cham. https://doi.org/10.1007/978-3-030-12524-0_11

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*(2), 101–121. https://doi.org/10.1006/jesp.1996.1314

Malle, B. F. (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. In *Advances in Experimental Social Psychology* (Vol. 44). Elsevier Inc. https://doi.org/10.1016/B978-0-12-385522-0.00006-8

Massin, O. (2017). Towards a definition of efforts. *Motivation Science, 3*(3), 230–259. https://doi.org/10.1037/mot0000066

Mastria, S., Agnoli, S., & Corazza, G. E. (2019). How does emotion influence the creativity evaluation of exogenous alternative ideas? *PLoS ONE, 14*(7), 1–16. https://doi.org/10.1371/journal.pone.0219298

Moffat, D. C., & Kelly, M. (2006). An investigation into people's bias against computational creativity in music composition. *Proceedings of the Third Joint Workshop on Computational Creativity*.

Mohr, L. A., & Bitner, M. J. (1995). The role of employee effort in satisfaction with service transactions. *Journal of Business Research, 32*(3), 239–252. https://doi.org/10.1016/0148-2963(94)00049-K

Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N., Davis, T., Waugh, K., Johanson, M., & Bowling, M. (2017). DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science, 513*(6337), 1–32. https://doi.org/10.1126/science.aam6960.1

Mueller, J. S., Melwani, S., & Goncalo, J. A. (2012). The bias against creativity: Why people desire but reject creative ideas. *Psychological Science, 23*(1), 13–17. https://doi.org/10.1177/0956797611421018

Mueller, J. S., Wakslak, C. J., & Krishnan, V. (2014). Construing creativity: The how and why of recognizing creative ideas. *Journal of Experimental Social Psychology, 51*, 81–87. https://doi.org/10.1016/j.jesp.2013.11.007

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Newman, G. E., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General, 141*(3), 558–569. https://doi.org/10.1037/a0026035

Newton, D. (2019). Generative deep learning in architectural design. *Technology|Architecture + Design, 3*(2), 176–189. https://doi.org/10.1080/24751448.2019.1640536

Ng, T. W. H., & Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology, 93*(2), 392–423. https://doi.org/10.1037/0021-9010.93.2.392

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283*.

Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal, 39*(3), 607–634. https://doi.org/10.2307/256657

Oliveira, H. G. (2012). PoeTryMe: A versatile platform for poetry generation. *ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence*.

Paciotti, D., & Di Stefano, A. (2021). Generative design and industrial product. Connecting physical/digital dimensions of the project. *AGATHÓN| International Journal of Architecture. Art and Design, 10*, 158–167.

Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research, 8*(2), 223–230. https://doi.org/10.1086/208859

Paulus, P. B., & Yang, H. C. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes, 82*(1), 76–87. https://doi.org/10.1006/obhd.2000.2888

Perry-Smith, J. E., & Mannucci, P. V. (2017). From creativity to innovation: The social network drivers of the four phases of the idea journey. *Academy of Management Review, 42*(1), 53–79. https://doi.org/10.5465/amr.2014.0462

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology, 66*(1), 29–38. https://doi.org/10.1016/j.jesp.2015.09.012

Popova, M., Isayev, O., & Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science Advances, 4*(7), aap7885. https://doi.org/10.1126/sciadv.aap7885

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting, 36*(6), 691–702. https://doi.org/10.1002/for.2464

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*(2), 77–98. https://doi.org/10.1080/19312458.2012.679848

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places. *Cambridge University Press*. https://doi.org/10.1109/MSPEC.1997.576013

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2006). Productivity is not enough: A comparison of interactive and nominal brainstorming groups on idea generation and selection. *Journal of Experimental Social Psychology, 42*(2), 244–251. https://doi.org/10.1016/j.jesp.2005.04.005

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2010). The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British Journal of Psychology, 101*(1), 47–68. https://doi.org/10.1348/000712609X414204

Rietzschel, E. F., Zacher, H., & Stroebe, W. (2016). A lifespan perspective on creativity and innovation at work. *Work, Aging and Retirement, 2*(2), 105–129. https://doi.org/10.1093/workar/waw005

Ritter, S. M., & Rietzschel, E. F. (2017). Lay theories of creativity. In C. M. Zedelius, B. C. N. Müller, & J. W. Schooler (Eds.),

The science of lay theories: How beliefs shape our cognition, behavior, and health. Springer International Publishing. https://doi.org/10.1007/978-3-319-57306-9

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*(6), 359–371. https://doi.org/10.1111/j.1751-9004.2011.00355.x

Runco, M. A. (2004). Creativity. *Annual Review of Psychology, 55*(1), 657–687. https://doi.org/10.1146/annurev.psych.55.090902.141502

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Runco, M. A., & Smith, W. R. (1992). Interpersonal and intrapersonal evaluations of creative ideas. *Personality and Individual Differences, 13*(3), 295–302. https://doi.org/10.1016/0191-8869(92)90105-X

Schmelzer, R. (2019). *Should we be afraid of AI?* Forbes. https://www.forbes.com/sites/cognitiveworld/2019/10/31/should-we-be-afraid-of-ai/. Accessed 29 Jun 2022.

Simonton, D. K. (2004). Creativity in science: Chance, logic, genius, and zeitgeist. *Cambridge University Press*. https://doi.org/10.1017/CBO9781139165358

Simonton, D. K. (2014). More method in the mad-genius controversy: A historiometric study of 204 historic creators. *Psychology of Aesthetics, Creativity, and the Arts, 8*(1), 53–61. https://doi.org/10.1037/a0035367

Sullivan, D. M., & Ford, C. M. (2010). The alignment of measures and constructs in organizational research: The case of testing measurement models of creativity. *Journal of Business and Psychology, 25*(3), 505–521. https://doi.org/10.1007/s10869-009-9147-8

Thellman, S., Silvervarg, A., & Ziemke, T. (2017). Folk-psychological interpretation of human vs humanoid robot behavior: Exploring the intentional stance toward robots. *Frontiers in Psychology, 8*(1962), 1–14. https://doi.org/10.3389/fpsyg.2017.01962

Tsay, C. J. (2016). Privileging naturals over strivers: The costs of the naturalness bias. *Personality and Social Psychology Bulletin, 42*(1), 40–53. https://doi.org/10.1177/0146167215611638

van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior, 90*, 215–222. https://doi.org/10.1016/j.chb.2018.09.009

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications, 11*(1), 233.

von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior, 26*(6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries, 4*(4), 404–409. https://doi.org/10.5465/amd.2018.0084

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206. https://doi.org/10.1086/651257

Zhou, J., Wang, X. M., Bavato, D., Tasselli, S., & Wu, J. (2019). Understanding the receiving side of creativity: A multidisciplinary review and implications for management research. *Journal of Management, 45*(6), 2570–2595. https://doi.org/10.1177/0149206319827088