Check for updates

The Journal of Creative Behavior

ORIGINAL ARTICLE OPEN ACCESS

Human-AI Co-Creativity: Does ChatGPT Make Us More Creative?

Kelsey Medeiros[1] | David H. Cropley[2,3] | Rebecca L. Marrone[3] | Roni Reiter-Palmon[4]

[1]Management Department, University of Nebraska at Omaha, Omaha, Nebraska, USA | [2]UniSA STEM, University of South Australia, Adelaide, South Australia, Australia | [3]Centre for Change and Complexity in Learning (C3L), University of South Australia, Adelaide, South Australia, Australia | [4]Psychology Department, University of Nebraska at Omaha, Omaha, Nebraska, USA

Correspondence: David H. Cropley (david.cropley@unisa.edu.au)

Received: 15 October 2024 | **Revised:** 2 April 2025 | **Accepted:** 7 April 2025

Funding: The authors received no specific funding for this work.

Keywords: ChatGPT | co-creativity | creativity | divergent association task | generative AI | idea generation

ABSTRACT

Much has been made of the apparent capacity for creativity of generative AI. However, as research expands the knowledge base regarding the capabilities and performance of this technology, the prevailing view is shifting away from "AI is creative" and towards a more balanced model of Human-AI co-creativity. Nevertheless, even this paradigm may be impacted by untested assumptions: for example, that generative AI will boost human performance on idea generation tasks. To test that assumption this study primed subjects with lists of words purportedly either from a human or from ChatGPT, and of varying degrees of creativity. Subjects then completed the Divergent Association Task (DAT). The results of this study found no evidence of any difference in divergent thinking resulting from either the source of priming (Human/ChatGPT) or from the level of creativity of the priming (low, typical, and high), *with one exception: a low-creativity prime, believed to be from ChatGPT*, resulted in lower scores on the DAT compared to other priming conditions. A subsequent regression analysis supported this result, finding only the perceived creativity of the prime to be a weak predictor of DAT scores (in addition to the expected trait of Openness). The consequences of these findings for Human-AI co-creativity are discussed.

1 | Introduction

Much has been made, since the release of widely accessible Large Language Models (LLMs), of the apparent creativity of generative AI. Early research in the period after the release of OpenAI's GPT-3, from approximately 2022, tended to claim that *AI is creative* or that it is *more creative than humans*. For example, Stevenson et al. (2022) state that LLMs produce "amazing written works..." (p. 1) and that these "meet the standard of the standard definition of creativity..." (p. 1). The same authors gave GPT-3 a variant of the well-known Alternate Uses Task (AUT), first developed by Guilford (1967), noting that, while humans generally performed better than GPT-3 on this task, it "is only a matter of time before GPT-3 catches up..." (p. 1). Guzik et al. (2023) went further, giving GPT-4 the Torrance Tests of Creative Thinking (TTCT: Torrance 1999), and finding that GPT-4 ranked in the top percentile for originality and fluency. They concluded, unequivocally, that *AI is creative*. Hubert et al. (2023) similarly operate on the premise that LLMs such as ChatGPT are "more creative than humans..." (p. 1). Chen and Ding (2023), meanwhile, conclude that GPT-4, on an "objective measure of creativity" (the Divergent Association Task (DAT): Olson et al. 2021), "outperforms 96% of humans" (p. 1).

Other research, however, has found conflicting results. Cropley (2023), for instance, looked beyond superficial results comparing GPT-4 mean scores on the DAT with a human baseline. That study discussed the implications of the

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *The Journal of Creative Behavior* published by Wiley Periodicals LLC on behalf of Creative Education Foundation (CEF).

variance in GPT-4 (percentile rank) results—ranging from 31.03 to 99.60—to explain that GPT-4, while outperforming the human sample on mean scores, nevertheless was an *unreliable* source of divergent ideation compared to any given human. Koivisto and Grassini (2023) similarly note that the best-performing humans on an AUT still "matched or exceed those of the chatbots" (p.1).

Still others, such as Giannuzzo (2023), approach the question of AI creativity from a more philosophical starting point. The capacity (or not) for creativity in AI is not a matter of the ability to replicate the process–divergent thinking–that is central to creativity, but a question of *why* that process is employed in the first place. Like Cropley et al. (2022), this argues that creativity cannot be separated from factors such as intent and autonomy, regardless of the efficiency of the core process.

At the same time, studies emerged that focused on *creativity as the capacity to produce creative outcomes* (a focus on the *end* of creativity) as opposed to *creativity as the capacity to execute a specific kind of divergent thinking* (a focus on the *means* of creativity). For instance, Chakrabarty et al. (2024) examined the capacity of LLMs to produce acclaimed creative writing. That study compared short stories written by humans, and published in the New Yorker magazine, with similar stories generated by Claude, GPT-3.5 and GPT-4. These stories were assessed by human, expert judges, using a detailed creativity rubric, without knowledge of the source of the stories (human or AI). Notably, these judges had an unerring ability to correctly detect the AI-written stories, and, found that the most creative stories were those written by humans. Critically, the judges found that the creative writing abilities of the AIs to be comparable to those only of *amateur* human writers. The authors additionally found that AI struggled with many important features of highly acclaimed creative writing (e.g., struggling with narrative ending, having poor rhetorical complexity, poor character development and weaknesses in narrative pacing). They concluded that, with respect to creativity, "there exists a wide chasm between them [LLMs] and human expertise when assessed holistically" (p. 18).

Some authors (e.g., Yiu et al. 2023) have been more forthright in their analysis. In a study that compared various LLMs and children with respect to both the capacity for *imitation* and for *innovation*, the authors concluded that the LLMs could not match the innovative capacity of a group of children with a mean age of 5.71 years.

Research is now increasingly turning to examine the potential for productive Human-AI co-creativity. Doshi and Hauser (2024), for example, examined the use of GPT-4 as an *aid* to creative writing. That study concluded that writers with lower levels of creative potential (as measured by their performance on the DAT) benefitted from using generative AI as an aid. Indeed, this lifted the creativity of their stories to a level comparable to writers with higher levels of creative potential. However, as the authors note, this rise in the creativity of outputs of individuals comes at the cost of collective novelty: the greater use of AI risks an overall homogenization (an increasing *sameness*) of creative writing. Niloy et al. (2024), in comparison, found that the use of ChatGPT for creative writing among university students was uniformly detrimental to creativity, while Rane et al. (2023), studying the use of ChatGPT for scientific writing, argued that LLMs cannot function as authors due to a lack of creativity, but that they have potential with regard to other key tasks such as assisting in data analysis and literature review. Finally, Jia et al. (2024) evaluated the use of AI for creative work performance and found that AI was beneficial for those that already were high performers but had no impact on lower performers. Consequently, whether AI can help all or some (and which ones), or hinder creative performance, is still unclear.

The present effort therefore focuses on this latter framework with the goal of shedding further light on human-AI co-creativity. Specifically, we examine differences in creative output based on the type of co-creator (human or ChatGPT) and the level of creative input (low, typical, or high) from the co-creator. Given the nascent nature of this body of work, our study uses a domain-general, verbal divergent ideation task (the DAT) as an indicator of, and proxy for, creativity (Olson et al. 2021). Consequently, conclusions from this study provide *initial* insights into the impact of the source and level of creativity of a co-creator on a necessary but not sufficient creative processing activity: divergent thinking (Runco and Acar 2019).

2 | The Impact of Co-Creator Characteristics on Divergent Thinking

2.1 | Source of Assistance

Although formal research examining differences in the type of co-creator is only beginning to emerge, work that has been conducted provides initial insights into how humans respond to human and AI co-creators. For example, Hwang and Won (2021) assessed how the perceived identity of a co-creator—human or chatbot—influenced creative outcomes. In a comparison of a brainstorming session, the study found that participants who worked with a bot produced more and lengthier ideas that were rated as more original compared to those participants who worked with a human. The researchers argued that this could be due to the social facilitation effects of working with a bot instead of a human or the novelty attached to working with new technology such as a bot.

Other work that has compared human vs. AI agents such as chatbots and robots suggests a less clear pattern. For example, Powers et al. (2007) compared reactions to agents, co-located robots, and remote robots, finding that participants were less likely to disclose information to a co-located robot compared to others. Given the importance of information sharing for creativity (Mesmer-Magnus and DeChurch 2009), this may be detrimental to creative performance. More recent work by Lee and Chung (2024) compared creative outcomes for those who used ChatGPT or a general web search. This work found that using ChatGPT resulted in more creative outcomes. Happle (2024) found that when scriptwriters use ChatGPT, there is a shift in agency and their roles, and that the impact of ChatGPT's assistance is largely dependent on the quality and nature of the prompts provided to it.

This research is valuable to our understanding of how the use of AI technologies influences creative outcomes. However, these

21626057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

influences unfold within complex and dynamic conditions, mirroring the realities of creative work. The benefit to this is, of course, that it represents the way creative activities unfold in the "real world." Under these conditions, however, it is difficult to isolate the unique impacts of particular variables on specific outcomes. Given the incipient nature of research on AI and creativity, it is simultaneously important to ask more fundamental, or basic, questions to support or eliminate explanations. In this study, we take a more controlled approach by reducing some of the complexities present in prior work. Specifically, we investigate how the source of creative examples—whether human or AI (ChatGPT)—affects the divergent thinking process. We have chosen ChatGPT due to its widespread accessibility and popularity, particularly when compared to embodied AI systems such as robots. Given the potential for perception of who the co-creator is to influence judgment regarding the ideas and collaboration (Pornpitakpan 2004), it is important to distinguish between the actual information provided and the perception of the collaborator.

2.2 | Level of Creativity of the Co-Creator

In addition to who assistance comes from, the level of creativity of that assistance may also influence subsequent creative output. Regarding AI, Haase and Hanel (2023) provide insight into how Human-AI co-creativity might be best exploited, drawing on the concept of the 4-Cs model of creativity (Kaufman and Beghetto 2009). They suggest that the greatest potential for co-creativity lies in *everyday* creativity, echoing the findings of Doshi and Hauser (2024) and Chakrabarty et al. (2024). In other words, the benefits of co-creativity—whether in terms of divergent ideation (the means) or the creative output (the end)—may depend on the *relative* abilities of the human and the AI. There is evidence (e.g., Chakrabarty et al. 2024) to support the notion that LLMs represent, at best, *typical* human ability, both in general (e.g., divergent ideation) and in domain-specific creative production (e.g., creative writing). Consequently, it is possible to hypothesize that a Pro-c individual (e.g., a professional creative writer) will find less value to be gained from the assistance of a mini-or little-c co-creator (e.g., an *amateur* LLM).

Therefore, with respect to creativity, it seems likely that, aside from speed, LLMs are of greatest benefit to those who are, relatively speaking, *amateurs*. However, even the purely quantitative benefits of Human-AI co-creativity are not to be overlooked. Zhou and Lee (2023) found that generative AI not only increased human creative productivity (in art) but that this increased the rate of favorable evaluations of the creative (artistic) outputs. Thus, even for above-average, or Pro-c/Big-C, individuals, there may be a creative advantage to be gained simply by being able to *think faster*.

The picture that emerges from the limited available literature is unclear. The rapid growth of LLMs in many areas of society, from education (Zhai et al. 2021) to professional practice (Chuma and De Oliveira 2023), warrants further investigation. Is human creativity enhanced by LLM assistance? Does the quality (i.e., the level of creativity) of the assistance influence the creative outcomes? To address these questions, we focus on a necessary but not sufficient component of creativity—Divergent Thinking—and examine the impact of human versus AI assistance on this capability. This leads to our first hypothesis:

Hypothesis 1. *There is no difference in the creativity* (*DAT*) *of participants based on the perceived source* (*Human or ChatGPT*) *and actual level (Low, Typical, High) of assistance given* (*the Prime*).

In addition, to better understand how the assistance given affects creative performance, we examine the individual's *perception* of the creativity of any assistance received (as distinct from its actual level). Do individuals, for example, perceive a high creativity prime as high in creativity? Given the lack of research in this area, we propose the following hypothesis:

Hypothesis 2. *There is no difference in the assessed creativity of the assistance given based on the perceived source* (*Human or ChatGPT*) *and actual level* (*Low, Typical, High*) *of the assistance.*

Finally, we recognize that other factors play a role in creativity, including, for example, personality (Kaufman et al. 2013). Therefore, we also examine the possible impact of other factors relevant to creativity (e.g., Openness to Experience or Creative Self Efficacy) within the context of co-creative assistance.

Hypothesis 3. *The kind of assistance given* (*source/level*) *predicts the creativity* (*DAT*) *of participants over and above other creativity variables.*

3 | Method

3.1 | Design and Participants

Participants ($N = 387$), drawn from an urban university in the Midwest of the United States, took part in a study using a 2 (Prime *Source:* Human vs. ChatGPT) x 3 (Prime Creativity: Low, Typical, High) factorial design. Participants responded to a set of demographic questions and a suite of personality, creativity, and AI covariates before being randomly allocated to one of six priming conditions. Participants rated the perceived creativity of the prime they received (a list of seven nouns) before completing the Divergent Association Task (DAT: see the *Measures* section).

Of the 387 participants, 107 identified as male, 273 identified as female, and 6 identified as non-binary, fluid, or transgender, while a single participant declined to respond. The majority (68.99%) of participants reported being white or Caucasian, and 4.13% reported being African or African American, 11.88% Hispanic Origin, 4.13% Asian or Asian American, 0.25% Native American, and 10.07% mixed race. The mean age was 23.87. The random allocation to priming conditions resulted in the following groups (Table 1).

3.2 | Procedure

Participants were first asked to provide demographic information including age, gender, and race. Next, participants completed a battery of covariates including measures of: (a) personality, (b)

21636057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

creative self-efficacy, (c) propensity to trust artificial intelligence, (d) creative personal identity, and (e) AI familiarity. Participants were then randomly allocated to one of six priming conditions based on: (a) the source of the prime (Human or ChatGPT) and (b) the level of creativity of the prime (Low, Typical, or High). Following this, participants were presented with the prime: an example of a Divergent Association Task (DAT) response (a list of seven words). To ensure that the participants had read this prime, they were asked to retype the list of words presented. Participants were then asked, "how creative do you think this list is?" Lastly, participants completed the DAT themselves. Figure 1 sets out the procedure for this factorial design.

3.3 | Priming

To explore the possible impact of an AI assistant (ChatGPT), compared to a human assistant, on divergent ideation, participants were allocated to a priming condition characterized by (a) the source (human or ChatGPT) and (b) the creativity level (low, typical or high) of a list of seven nouns.

After random allocation to the priming condition, participants were presented with a list of seven nouns and were told that these came from either "a student" or from "ChatGPT." However, they were *not* told the level of creativity (low, typical or high) of the list of seven nouns (as measured on the DAT). Thus, participants in the "low" creativity condition (regardless of source) received a unique list of nouns (Table 2); participants in the "typical" creativity condition (regardless of source) received a different and unique list of nouns; and the participants in the "high" creativity condition received a third unique list of nouns. The creativity level of the list of nouns was determined from a sample (Cropley 2023) of ChatGPT responses ($N = 102$) with a mean Percentile Rank (relative to human norms) of 64.93 (SD = 21.75). The *low* creativity response (the list of seven nouns) was selected with a PR = 0.05, the *typical* creativity response list was selected with a PR = 64.69, and the *high* creativity response list was selected with a PR = 95.75 (see Table 2). Note that the *typical* creativity list was selected to be as close to the mean of ChatGPT responses as possible.

TABLE 1 | Priming conditions (source, creativity).

Source	Creativity	*n*
Human	Low	56
	Typical	69
	High	67
ChatGPT	Low	53
	Typical	74
	High	68
Total (*N*)		387

3.4 | Attention Check

At the conclusion of the study, participants were also asked to complete an attention check. This asked participants to select a particular answer based on the text of the attention check. Participants who selected the incorrect answer were removed from the sample.

3.5 | Measures

3.5.1 | Covariates

3.5.1.1 | **Personality.** Soto and John's (2017) 60-item measure was used to assess participants' openness to experience. Sample items include "I am someone who...": "... is outgoing, sociable", "...tends to be lazy", and "...is dominant, acts as a leader." Soto and John (2017) reported reliability coefficients from $\alpha = 0.83$. In the present study, reliability coefficients ranged from $\alpha = 0.77$ to $\alpha = 0.88$. Of particular importance to this study, the reliability coefficient for the openness subscale was acceptable ($\alpha = 0.82$).

3.5.1.2 | **Creative Self Efficacy.** Tierney and Farmer's (2002) 3-item measure was used to assess creative self-efficacy. The

TABLE 2 | Priming lists.

Low creativity (PR = 0.03)	Typical creativity (PR = 64.69)	High creativity (PR = 95.75)
Chair	Phoenix	Hopper
Table	Equinox	Torpedo
Window	Labyrinth	Omelette
Door	Chimera	Brother
Floor	Elixer	Thumb
House	Mirage	Escapement
Carpet	Tornado	Neutrino

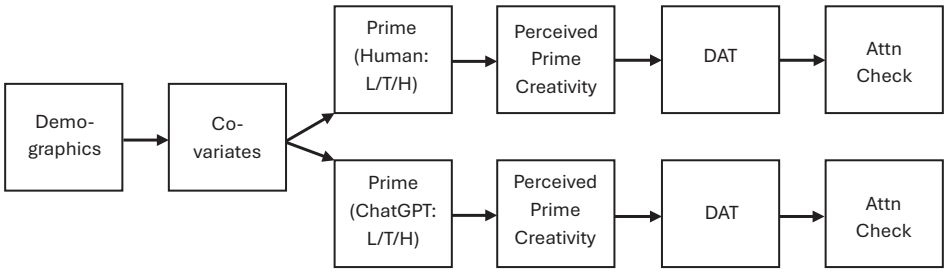


FIGURE 1 | Study procedure.

21626057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

items were: "I am good at coming up with new ideas", "I have a lot of good ideas", and "I have a good imagination." Tierney and Farmer (2002) reported a reliability coefficient of $\alpha = 0.83$. In the present study, $\alpha = 0.82$.

3.5.1.3 | Propensity to Trust. Merritt's (2011) six-item measure was adapted to measure participants' propensity to trust artificial intelligence. Sample items include: "I usually trust artificial intelligence (AI) until there is a reason not to", "In general, I would rely on an artificial intelligence (AI) to assist me", and "I am likely to trust an artificial intelligence (AI) even when I have little knowledge of it." Merritt (2011) reported a reliability coefficient of $\alpha = 0.87$. In the present study, $\alpha = 0.65$.

3.5.1.4 | Creative Personality Identity. Farmer et al. (2003) three-item measure was used to assess participants' creative personality identity. The items were: "I often think about being creative", "I do not have any clear concept of myself as a creative person", and "to be a creative person is an important part of my identity." Farmer et al. (2003) reported a reliability coefficient of $\alpha = 0.80$. In the present study, $\alpha = 0.77$.

3.5.1.5 | AI Familiarity. Participants were also asked to respond to two items regarding their recent exposure to artificial intelligence and ChatGPT. Specifically, participants were asked, "Have you read anything about artificial intelligence (AI) recently?" and "Have you read anything about ChatGPT recently?"

3.5.2 | Dependent Variables

3.5.2.1 | Perceived Creativity of the Prime. To rate the perceived creativity of the prime that they received (the list of seven nouns), participants were presented with a five-point, Likert-type scale, ranging from 1 (Uncreative) through 3 (Neither creative nor uncreative) to 5 (Creative).

3.5.2.2 | Divergent Association Task. The Divergent Association Task (DAT: Olson et al. 2021) was used to assess participants' verbal divergent ideation. The DAT instructions (p. 5) invite participants to do the following:

Please enter 10 words that are as different from each other as possible, in all meanings and uses of the words.

Rules:

1. Only single words in English.
2. Only nouns (e.g., things, objects, concepts).
3. No proper nouns (e.g., no specific people or places).
4. No specialized vocabulary (e.g., no technical terms).
5. Think of the words on your own (e.g., do not just look at objects in your surroundings).

The DAT has shown strong convergent validity with the Alternate Uses Task, with statistically significant, positive correlations as high as $r = 0.51$ for flexibility and $r = 0.50$ for originality (Olson et al. 2021). Other studies using the same underlying method (semantic distance) have found similar positive correlations, both to the AUT and to creative personality variables (e.g., Beaty et al. 2022).

4 | Results

Preliminary graphical and numerical assessments of skewness and kurtosis indicated that the data were normally distributed, and therefore suitable for parametric analysis. Table 3 lists basic descriptive statistics for the variables under investigation.

4.1 | The Impact of Priming on Verbal Divergent Ideation (H1)

A one-way between-groups analysis of variance (ANOVA) was conducted to explore the impact of the priming condition on creativity (verbal divergent ideation) as measured by the Divergent Association Task (DAT). Participants were divided into six groups according to the priming source (Human or ChatGPT) and level of creativity (Low, Typical, High). These groups were designated as HuLo: Human/Low; HuTy: Human/Typical; HuHi: Human/High; ChLo: ChatGPT/Low; ChTy: ChatGPT/Typical; ChHi: ChatGPT/High. There was a statistically significant difference in

TABLE 3 | Descriptive Statistics.

Variable	*N*	Minimum	Maximum	Mean	SD
Openness	384	2.00	5.00	3.72	0.63
Propensity to trust AI	384	1.50	5.00	2.85	0.66
Creative self-efficacy	387	1.00	5.00	3.86	0.75
Creative personal identity	387	1.00	5.00	3.47	0.87
Prime creativity	387	1.00	5.00	3.42	1.32
DAT score	381	61.16	94.45	80.25	6.30
DAT percentile rank	381	0.59	99.31	60.28	27.59
Total (valid)	376				

21636057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DAT scores for the six priming condition groups: $F(5, 375) = 6.39$, $p < 0.001$. The actual difference in mean scores between the groups was moderate, with the effect size, calculated using eta squared, = 0.08.

Post hoc comparisons using the Tukey HSD test indicated that the mean DAT score for the group ChLo ($M = 76.37$, SD = 5.71) was significantly different from all other conditions (HuLo: $M = 79.96$, SD = 5.94; HuTy: $M = 79.74$, SD = 7.24; HuHi: $M = 82.18$, SD = 5.63, ChTy: $M = 81.21$, SD = 6.07; ChHi: $M = 81.11$, SD = 5.66). Consequently, H1 was not supported: there is, in fact, some difference in participant DAT scores attributable to the priming condition (see Figure 2).

4.2 | The Assessed Creativity of the Prime (H2)

A one-way between-groups analysis of variance (ANOVA) was conducted to explore the impact of the priming condition on perceived response creativity as measured on a 5-pt Likert-type scale (1 = *uncreative* to 5 = *creative*). Participants were divided into six groups according to the priming source (Human or ChatGPT) and level of creativity (Low, Typical, High). These groups were designated as HuLo: Human/Low; HuTy: Human/Typical; HuHi: Human/High; ChLo: ChatGPT/Low; ChTy: ChatGPT/Typical; ChHi: ChatGPT/High. There was a statistically significant difference in perceived response creativity scores for the six priming condition groups: $F(5, 381) = 71.90$, $p < 0.001$. The actual difference in mean scores between the groups was very large, with the effect size, calculated using eta squared, = 0.48.

Post hoc comparisons using the Tukey HSD test indicated that the mean scores for the group HuLo ($M = 1.91$, SD = 0.98) and ChLo ($M = 2.04$, SD = 1.14) were significantly different from HuTy ($M = 4.16$, SD = 0.85), HuHi ($M = 4.13$, SD = 0.80), ChTy ($M = 3.73$, SD = 1.01) and ChHi ($M = 3.93$, SD = 0.94). Therefore, H2 is not supported: there appears to be a difference in the perceived response creativity of the prime attributable to the source and/or actual level of creativity of the prime (see Figure 3). Specifically, participants were able to recognize that the low creativity prompts were indeed lower in creativity but were not able to distinguish between typical and high creativity.

4.3 | Does the Kind of Prime Predict Verbal Divergent Ideation (H3)?

Hierarchical multiple regression was used to explore the ability of openness, creative self-efficacy (CSE), creative personal identity (CPI), trust in AI (TAI), perceived response creativity, and priming condition to predict levels of creativity (verbal divergent production: DAT). Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, multicollinearity, and homoscedasticity. Specifically, and consistent with H2, the VIF statistic indicated no significant or severe multicollinearity between the perceived creativity of the prime and the actual prime creativity (VIF = 1.063 for both variables in the regression analysis). Openness was entered at step 1, with the model significantly predicting creativity (DAT), $F(1, 376) = 19.87$, $p < 0.001$, explaining 5% of the variance in creativity (DAT). After entry of CSE and CPI at step 2, the model as a whole explained 5.7% of the total variance, $F(3, 374) = 7.47$, $p < 0.001$; however, R squared change = 0.006, F change $(2, 374) = 1.26$, $p = 0.285$. Next, after entry of TAI at step 3, the model as a whole explained 5.8% of the total variance, $F(4, 373) = 5.74$, $p < 0.001$; however, R squared change = 0.001, F change $(1, 373) = 0.57$, $p = 0.452$. Next, after entry of perceived response creativity in step 4, the model as a whole explained 8.7% of the total variance, $F(5, 372) = 7.08$, $p < 0.001$, with R squared change = 0.029, F change $(1, 372) = 11.81$, $p < 0.001$. Finally, after entry of the priming condition at step 5, the model as a whole explained 8.7% of the total variance, $F(6, 371) = 5.90$, $p < 0.001$; however, R squared change = 0.000, F change $(1, 371) = 0.05$, $p = 0.824$. Consequently, H3 is not supported: it appears that the priming, as a combination of source (Human or ChatGPT) and actual level of creativity (Low, Typical, High), does not predict verbal divergent ideation above and beyond other variables.

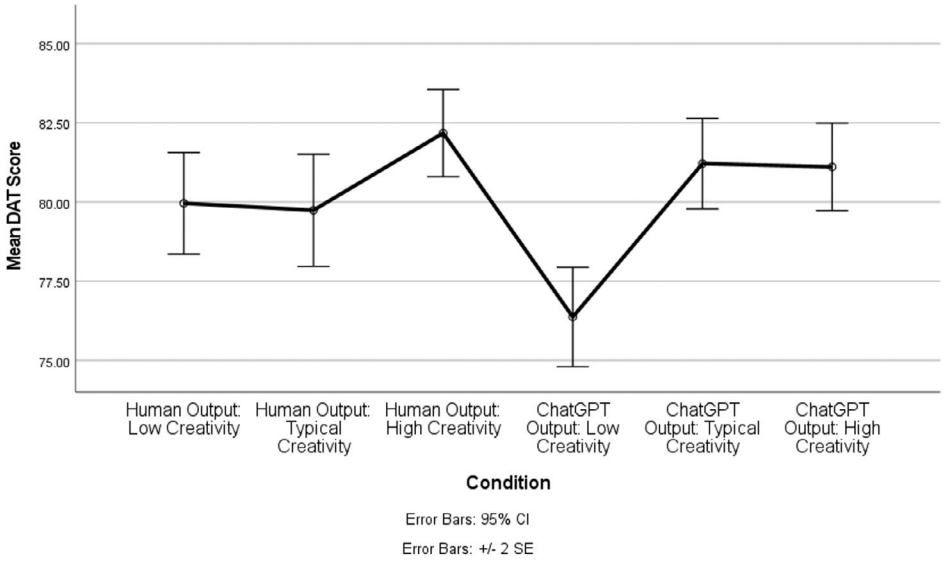


FIGURE 2 | Priming condition and DAT scores.

21626057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

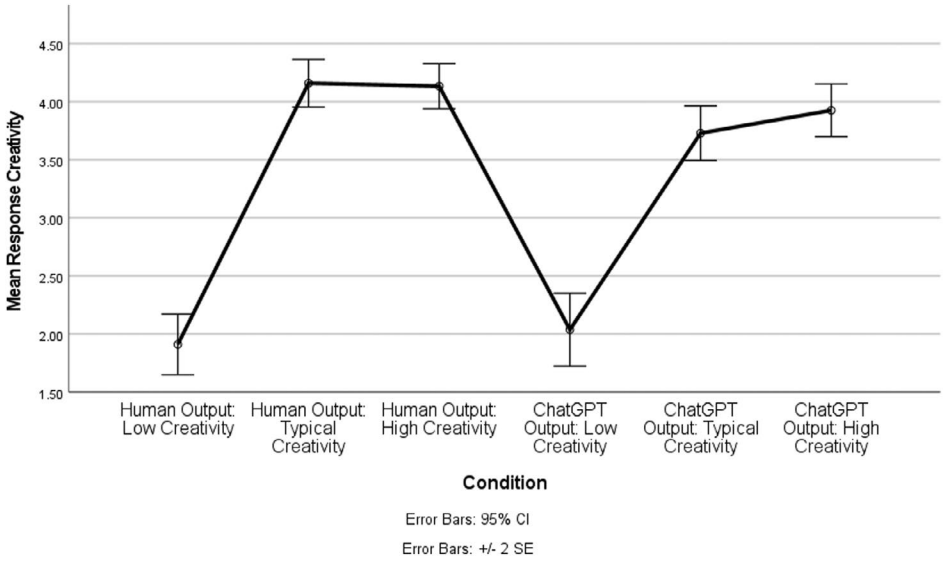


FIGURE 3 | Priming condition and perceived response creativity of the prime.

5 | Discussion

As both popular and scholarly debate regarding creativity and generative AI shifts away from absolutist claims that AI is creative (and therefore will replace human creativity) and towards a more nuanced model of Human-AI co-creativity, there remain many unanswered questions. While it is clear that generative AI performs many of the tasks that are building blocks of creativity—divergent ideation, for example—far more quickly than humans, and frequently at levels that are, on average, superior to human responses, far less is known about the impact of integrating this capability into human creative problem-solving processes (e.g., divergent ideation is only one aspect of the creative process; Marrone et al. 2024). The present study therefore set out to explore the practicalities of Human-AI co-creativity by focusing on the potential for generative AI to boost human performance on an idea generation task.

The first hypothesis in this study (H1) proposed that there would be no difference in the creativity (as measured on the Divergent Association Task (DAT)) of participants based on the source (Human/ChatGPT) or level of creativity (Low, Typical, High) of assistance given to the participants (in the form of a list of sample responses to the DAT). The data, however, indicated that there was one assistance condition—a low-creativity list of words generated by ChatGPT—that did, in fact, lead to a statistically significant reduction in the creativity of participants. Conversely, all other conditions resulted in higher (and statistically indistinguishable) creativity on the DAT. In the context of Human-AI co-creativity, these results therefore suggest that: (a) the qualitative benefit of AI assistance, over and above human assistance, may be negligible, and (b) low-quality AI assistance may, in fact, hinder human creativity. Thus, while generative AI may provide quantitative (i.e., speed or fluency) benefits to Human-AI co-creativity, care must be taken to avoid potential negative effects on quality. Indeed, Hwang and Won (2021) support the *quantitative* benefits of AI assistance (to ideational fluency and elaboration), with the present study suggesting that this may be moderated by the actual creativity of the AI assistance.

Why did low-creativity ChatGPT assistance hinder participant creativity when the same assistance from a human did not? One possible explanation for this is that participants perceived the creativity of the different sources (Human/ChatGPT) differently. The second hypothesis therefore sought to explore differences in perceived, as opposed to actual, creativity based on the source of the assistance. The data, in this case, indicated that participants correctly identified low-creativity assistance from both sources as low in creativity, whereas they identified all other assistance from both sources as high in creativity.

These results suggest that participants were particularly adept at identifying, or sensitive to, low-creativity assistance. Indeed, they may perceive the creativity of the assistance more simplistically as either "low" or "not low". Most importantly, however, is that even though participants know that the assistance is low creativity, they react quite differently to it depending on the perceived source. Why do participants, in effect, ignore the low-creativity assistance when it comes from a human, but are negatively impacted when the same assistance comes from ChatGPT?

A possible explanation for the observed results may be found in the concept of the *self-fulfilling prophecy* (see Romney et al. 2024) and more specifically, the *Golem* Effect. This effect posits that poorer performance (in this case, on the DAT) may result from participants internalizing a belief that the low-creativity, ChatGPT assistance they received is evidence of lower expectations that have been placed on this group. This may, indeed, be an example of what Davidson and Eden (2000) described as a *relative* Golem Effect, with even otherwise highly skilled (i.e., highly creative) individuals suffering degraded performance. Significantly, this *Golemization* appears to be unique to the low-creativity assistance provided by ChatGPT, suggesting that participants may tend to see

21626057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/joch.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

generative AI as infallible. Conversely, participants remain far more willing to challenge, and if necessary, dismiss or ignore, assistance provided by humans.

To understand more deeply the impact of the assistance given to participants (both source and level), this study concluded by exploring the ability of that assistance to predict DAT scores over and above other well-established variables (e.g., openness). These results merely confirm that openness is a significant predictor of creativity (and divergent thinking), while the perceived creativity of the assistance (in this case, the low-creativity prime) has some predictive power. Interestingly, given the hypothesized influence of the Golem Effect, *Trust in AI* was not a significant predictor of creativity. This may suggest that the negative impact of the low-creativity ChatGPT assistance is less a function of the quality of the AI assistance and more a function of how individuals feel about receiving assistance from a non-human source. Marmolejo-Ramos (2023) argues that trust in AI is shaped by statistical literacy, with more knowledgeable users being more critical of AI. This could imply that participants were not critical enough of ChatGPT's low-creativity outputs. However, these authors do not measure emotional responses to AI, which is crucial to understanding human choices. This area warrants further research as it could also explain our findings.

6 | Practical Implications

Practical implications of these findings include the need for careful design and implementation of AI systems used in creative contexts. For instance, ensuring that AI-generated prompts are of high quality can prevent the negative impact observed with low-creativity inputs. To address the negative impact of low-quality outputs, AI systems could be designed with filters or thresholds that ensure only sufficiently creative outputs are presented to users.

Additionally, the study's findings also underscore the importance of training users to critically evaluate AI-generated content. Users should be educated about the capabilities and limitations of AI, recognizing that while AI can assist in creative processes, it is not infallible and may produce subpar or uncreative outputs. By fostering an understanding that AI-generated content should be critically assessed, and not taken at face value, users can become more discerning in their use of AI tools. This critical evaluation can mitigate the potential negative effects of low-creativity AI prompts and help maintain high standards of creativity. In educational settings, this implies the need for careful selection and quality control of AI-generated content to ensure that students are exposed to high-quality creative prompts, which can foster critical thinking and innovation. Educators should adequately train students to critically evaluate AI outputs and use AI as a tool to inspire, rather than as a crutch. In business, AI can streamline creative processes and provide inspiration, but companies must ensure that AI-generated ideas meet a certain standard to avoid demotivating employees. Businesses should implement practices that balance AI's efficiency with human oversight to maintain high levels of creativity and innovation. Lastly, both schools and businesses incorporating generative AI into creativity and creative processes must be aware of the expectations—both positive and negative—that this may create.

6.1 | Limitations

Although this work provides valuable insights into how generative AI may help, or in this case hinder, creative performance, there are several limitations that should be borne in mind when interpreting the results. First, this research was conducted using a university student sample, which may not be representative of the population at large. Second, this research was conducted at a single university in the Midwest of the United States. This limits the interpretation of results to similar groups, as the present effort does not account for cultural differences in perceptions of or experience with AI technology beyond that of a WEIRD sample. Third, the present effort explored one aspect of creativity—divergent thinking. Divergent thinking, while central to many definitions of creativity, represents only one facet of creativity. This is an important initial step in understanding a seemingly complex relationship, but caution should be taken when considering how these results inform creative performance more generally.

6.2 | Future Research

The present effort suggests several avenues for future research. Perhaps the clearest next step is to further examine the role individual differences and perceptions play in the interpretation of creative examples from different sources. Specifically, understanding how openness to experience influences how individuals view examples of varying levels of creativity from different sources would be particularly useful for informing team formation and who is best suited to work with AI on creative projects. Relatedly, it is important to understand the role of cultural differences and related factors on human interaction with AI in creative efforts.

Additionally, as the current effort only explored divergent thinking as an outcome, future research should explore other creativity-related processes and activities. For example, future research could explore how AI versus human prompts influence problem identification and idea evaluation. Expanding the scope of inquiry to include other phases of the creative process, such as incubation, elaboration, and verification, could provide a more holistic understanding of AI's impact on creativity. Further, examining the use of AI throughout a creative project lifecycle would be particularly informative. Exploring the use of AI in the field using qualitative methods would be illuminating, as would examining the same relationship empirically or qualitatively in a laboratory setting.

7 | Conclusion

As generative AI evolves and becomes more ubiquitous, research continues to examine the impact of this technology on the creative capacity of humans. Although Human-AI co-creativity is becoming accepted as the best approach, many questions remain to be answered. Studies continue to operate from a premise that AI is unequivocally beneficial to human creativity, but this assumption must be challenged and explored in more detail. The present study provides some evidence that care must be taken in a Human-AI interaction. Generative AI,

21626057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

as an assistant to human divergent ideation, while no better than a human assistant, is capable of dragging down elements of creative performance. In other words, while it may not offer any clear benefit, over and above human assistance, generative AI is capable of undermining creativity. It is not clear why this is the case, although it is possible to speculate that humans, while not necessarily overly trusting of generative AI, may, nevertheless, be unwilling to challenge the outputs of tools such as ChatGPT. The impact that this may have on Human-AI co-creativity requires significant further research.

Acknowledgments

Open access publishing facilitated by University of South Australia, as part of the Wiley - University of South Australia agreement via the Council of Australian University Librarians.

Ethics Statement

This research project was reviewed for ethics and approved by the Institutional Review Board at the University of Nebraska at Omaha (Application: 0223-23-EX).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data available on request.

References

Beaty, R. E., D. R. Johnson, D. C. Zeitlen, and B. Forthmann. 2022. "Semantic Distance and the Alternate Uses Task: Recommendations for Reliable Automated Assessment of Originality." *Creativity Research Journal* 34, no. 3: 245–260. https://doi.org/10.1080/10400419.2022.2025720.

Chakrabarty, T., P. Laban, D. Agarwal, S. Muresan, and C.-S. Wu. 2024. Art or Artifice? Large Language Models and the False Promise of Creativity Paper presented at the Proceedings of the CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3613904.3642731.

Chen, H., and N. Ding. 2023. *In Findings of the Association for Computational Linguistics: EMNLP 2023*, 12881–12888. Association for Computational Linguistics.

Chuma, E. L., and G. G. De Oliveira. 2023. "Generative AI for Business Decision-Making: A Case of ChatGPT." *Management Science and Business Decisions* 3, no. 1: 5–11.

Cropley, D. H. 2023. "Is Artificial Intelligence More Creative Than Humans?" *Learning Letters* 2: 13.

Cropley, D. H., K. Medeiros, and A. Damadzic. 2022. "Creativity and Artificial Intelligence: The Integration of Human and Artificial Cognition." In *Creative Provocations: Speculations on the Future of Creativity, Technology & Learning*, edited by D. Henriksen and P. Mishra, 19–34. Springer.

Davidson, O. B., and D. Eden. 2000. "Remedial Self-Fulfilling Prophecy: Two Field Experiments to Prevent Golem Effects Among Disadvantaged Women." *Journal of Applied Psychology* 85, no. 3: 386–398. https://doi.org/10.1037//0021-9010.85.3.386.

Doshi, A. R., and O. P. Hauser. 2024. "Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content." *Science Advances* 10, no. 28: adn5290. https://doi.org/10.1126/sciadv.adn5290.

Farmer, S., P. Tierney, and K. Kung-McIntyre. 2003. "Employee Creativity in Taiwan: An Application of Role Identity Theory." *Academy of Management Journal* 46: 618–630. https://doi.org/10.5465/30040653.

Giannuzzo, A. 2023. "Creativity, Intentions, and Self-Narratives: Can AI Really be Creative?" Paper presented at the EPIA Conference on Artificial Intelligence. https://doi.org/10.1007/978-3-031-49011-8_5.

Guilford, J. P. 1967. *The Nature of Human Intelligence*. McGraw-Hill.

Guzik, E. E., C. Byrge, and C. Gilde. 2023. "The Originality of Machines: AI Takes the Torrance Test." *Journal of Creativity* 33, no. 3: 100065.

Haase, J., and P. H. Hanel. 2023. "Artificial Muses: Generative Artificial Intelligence Chatbots Have Risen to Human-Level Creativity." *Journal of Creativity* 33, no. 3: 100066.

Happle, K. M. 2024. "Cre-AI-Tivity: The Intersection of Human Ingenuity and AI Collaboration in Scriptwriting With ChatGPT 4.0." (Master's thesis, Utrecht University). https://studenttheses.uu.nl/bitstream/handle/20.500.12932/46352/HappleK_1698397_MasterThesis.pdf?sequence=1.

Hubert, K., K. N. Awa, and D. Zabelina. 2023. "Artificial Intelligence is More Creative Than Humans: A Cognitive Science Perspective on the Current State of Generative Language Models."

Hwang, A. H. C., and A. S. Won. 2021. "IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity." In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.

Jia, N., X. Luo, Z. Fang, and C. Liao. 2024. "When and How Artificial Intelligence Augments Employee Creativity." *AMJ* 67: 5–32.

Kaufman, J. C., and R. A. Beghetto. 2009. "Beyond Big and Little: The Four-c Model of Creativity." *Review of General Psychology* 13: 1–12.

Kaufman, J. C., T. T. Pumaccahua, and R. E. Holt. 2013. "Personality and Creativity in Realistic, Investigative, Artistic, Social, and Enterprising College Majors." *Personality and Individual Differences* 54, no. 8: 913–917.

Koivisto, M., and S. Grassini. 2023. "Best Humans Still Outperform Artificial Intelligence in a Creative Divergent Thinking Task." *Scientific Reports* 13, no. 1: 13601. https://doi.org/10.1038/s41598-023-40858-3.

Lee, B. C., and J. J. Chung. 2024. "An Empirical Investigation of the Impact of ChatGPT on Creativity." *Nature Human Behaviour* 8: 1906–1914.

Marmolejo-Ramos, F. 2023. "Trust in Algorithms. An Experimental Approach." PsyArXiv. https://doi.org/10.31234/osf.io/9wh2f.

Marrone, R., D. H. Cropley, and K. E. Medeiros. 2024. "How Does Narrow AI Impact Human Creativity?" *Creativity Research Journal* 2: 1–11. https://doi.org/10.1080/10400419.2024.2378264.

Merritt, S. 2011. "Affective Processes in Human-Automation Interactions." *Human Factors* 53: 356–370. https://doi.org/10.1177/0018720811411912.

Mesmer-Magnus, J. R., and L. A. DeChurch. 2009. "Information Sharing and Team Performance: A Meta-Analysis." *Journal of Applied Psychology* 94, no. 2: 535–546.

Niloy, A. C., S. Akter, N. Sultana, J. Sultana, and S. I. U. Rahman. 2024. "Is ChatGPT a Menace for Creative Writing Ability? An Experiment." *Journal of Computer Assisted Learning* 40, no. 2: 919–930.

Olson, J. A., J. Nahas, D. Chmoulevitch, S. J. Cropper, and M. E. Webb. 2021. "Naming Unrelated Words Predicts Creativity." *Proceedings of the National Academy of Sciences* 118, no. 25: 2635.

Pornpitakpan, C. 2004. "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34: 243–281.

21636057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Powers, A., S. Kielser, S. Fussell, and C. Torrey. 2007. "Comparing a Computer Agent With a Humanoid Robot." Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction.

Rane, N. L., S. P. Choudhary, A. Tawde, and J. Rane. 2023. "ChatGPT Is Not Capable of Serving as an Author: Ethical Concerns and Challenges of Large Language Models in Education." *International Research Journal of Modernization in Engineering Technology and Science* 5, no. 10: 851–874. https://doi.org/10.56726/IRJMETS45212.

Romney, A., J. T. Harrison, and S. Benson. 2024. "Looking Back to Predict the Future: A Review of Empirical Support for the Self-Fulfilling Prophecy." *Management Research Review* 47, no. 2: 245–260.

Runco, M. A., and S. Acar. 2019. "Divergent Thinking." In *The Cambridge Handbook of Creativity*, edited by J. C. Kaufman and R. J. Sternberg, 2nd ed., 224–254. Cambridge University Press.

Soto, C. J., and O. P. John. 2017. "The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model With 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power." *Journal of Personality and Social Psychology* 113, no. 1: 117–143. https://doi.org/10.1037/pspp0000096.

Stevenson, C., I. Smal, M. Baas, R. Grasman, and H. van der Maas. 2022. Putting GPT-3's Creativity to the (Alternative Uses) Test arXiv:2206.08932.

Tierney, P., and S. M. Farmer. 2002. "Creative Self-Efficacy: Its Potential Antecedents and Relationship to Creative Performance." *Academy of Management Journal* 45, no. 6: 1137–1148. https://doi.org/10.5465/3069429.

Torrance, E. P. 1999. *Torrance Test of Creative Thinking: Norms and Technical Manual*. Scholastic Testing Services.

Yiu, E., E. Kosoy, and A. Gopnik. 2023. "Transmission Versus Truth, Imitation Versus Innovation: What Children Can Do That Large Language and Language-And-Vision Models Cannot (Yet)." *Perspectives on Psychological Science* 19, no. 5: 874–883.

Zhai, X., X. Chu, C. S. Chai, et al. 2021. "A Review of Artificial Intelligence (AI) in Education From 2010 to 2020." *Complexity* 2021, no. 1: 8812542. https://doi.org/10.1155/2021/8812542.

Zhou, E., and D. Lee. 2023. "Generative Ai, Human Creativity, and Art." SSRN. https://doi.org/10.1093/pnasnexus/pgae052.

21626057, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jocb.70022 by UNIVERSITY OF SOUTH AUSTRALIA, Wiley Online Library on [29/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License