ARTICLE IN PRESS

International Journal of Research in Marketing xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

# Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations

Begum Celiktutan [1], Anne-Kathrin Klesse [1,*], Mirjam A. Tuk [1]

*Rotterdam School of Management, Erasmus University, Rotterdam, 3062 PA, The Netherlands*

ARTICLE INFO

*Article history:*
Received 2 December 2023
Available online xxxx
Accepting Editor: Martin Schreier

*Keywords:*
GenAI
ChatGPT
Inferred contribution
Intellectual ownership
Self-other difference
Biased self-evaluation

ABSTRACT

Since the release of ChatGPT, heated discussions have focused on the acceptable uses of generative artificial intelligence (GenAI) in education, science, and business practices. A salient question in these debates pertains to perceptions of the extent to which creators contribute to the co-produced output. As the current research establishes, the answer to this question depends on the evaluation target. Nine studies (seven preregistered, total $N = 4498$) document that people evaluate their own contributions to co-produced outputs with ChatGPT as higher than those of others. This systematic self–other difference stems from differential inferences regarding types of GenAI usage behavior: People think that they predominantly use GenAI for inspiration, but others use it to outsource work. These self–other differences in turn have direct ramifications for GenAI acceptability perceptions, such that usage is considered more acceptable for the self than for others. The authors discuss the implications of these findings for science, education, and marketing.


## 1. Introduction

In the last decade, significant advances in artificial intelligence (AI) have revolutionized various aspects of our lives. Initially, AI functioned predominantly to enhance the decisions and processes adopted by businesses (Bonezzi & Ostinelli, 2021; Gai & Klesse, 2019), institutions (Cadario et al., 2021; Yalcin et al., 2023), and governments (De la Garza, 2020). More recently though, the development and introduction of generative AI (GenAI) tools, such as ChatGPT, Magic Write, and DALL-E2, have extended access to the general public. These tools can generate seemingly intelligent responses to human prompts, including texts, code, or images (Peres et al., 2023). In particular, ChatGPT (or "Chat Generative Pre-Trained Transformer") has vastly expanded the role of GenAI in society, largely because this free tool requires no prior knowledge or expertise, and it can assist with a broad range of tasks, ranging from the development of a marketing plan to writing scientific papers.

This rapid development and diffusion of GenAI in turn has raised a vast range of questions that deserve scientific attention, as spelled out by Peres et al. (2023). Most research into GenAI thus far has sought to understand its *objective performance*, such as how and when ChatGPT can increase productivity. Yet users' *subjective perceptions* of GenAI, including their sense of what it is capable of and whether its usage is acceptable, are pivotal for determining rules and legislations surrounding its usage. For example, many educators acknowledge that GenAI can be a valuable tool (e.g., Dwivedi et al., 2023) but also warn against over-reliance or outsourcing assignments completely to the technology (Roose, 2023). Similarly,

* Corresponding author.
*E-mail address:* klesse@rsm.nl (A.-K. Klesse).
[1] All authors contributed equally and are listed alphabetically.

https://doi.org/10.1016/j.ijresmar.2024.05.006
0167-8116/© 2024 The Authors. Published by Elsevier B.V.

scientific journals have established policies indicating that it is unacceptable to have ChatGPT co-author academic work (Peres et al., 2023; Thorp, 2023).

Such concerns may stem, at least partly, from the difficulty of judging the extent to which output has been produced by GenAI or humans. Beliefs about the extent to which an individual creator can be credited for such output may lie in the proverbial eye of the beholder, such that they vary depending on whether the judgment pertains to people's own work or the work of others. For example, academics determine how much *they* relied on GenAI when writing an academic article (i.e., own contribution) and also how much *others* used GenAI when reviewing articles (i.e., others' contribution). We predict that people's subjective perception of how much a person has contributed to a task when it was co-produced with GenAI (hereafter 'inferred contribution') is higher when individuals assess their *own* contribution as compared to when they infer *others*' contribution to the output. This prediction aligns with documented, systematic differences in how people assess themselves versus others in various domains (Epley & Dunning, 2000; Pronin et al., 2002; Williams et al., 2012). Specifically, people are more likely and more motivated to consider their own (positive) thoughts, feelings, and intentions when evaluating their own actions, but they only evaluate the observable behavior of others (Pronin et al., 2002). In turn, we posit that when evaluating their own behavior, people focus on their positive intentions, namely, using GenAI for inspiration (but not task completion). However, when evaluating others' behavior, they focus on the action, that is, completing a task with a tool that might have done all the work. Such differential inferences about usage behavior then should result in a self–other difference in inferred contributions.

We test these predictions with nine experiments ($N = 4498$) that feature nationally representative and convenience samples, real and imagined usage behavior, and different usage domains. The empirical findings consistently reveal a self–other difference in inferred contribution and further document the ramifications of this difference for people's evaluations of whether GenAI usage is acceptable or not.

Accordingly, this research makes several important contributions. First, it responds to calls for more research on GenAI and its implications for science, education, and business practices (e.g., Peres et al., 2023). While some of the queries raised relate to the *objective performance* of GenAI, others pertain to the *subjective assessment of performance,* such as the extent to which creators get credit for work created with GenAI. As we document, perceptions of inferred contributions to a task depend on whether this judgment refers to the self or others. By establishing that the extent to which a creator receives credit for an output co-produced with GenAI depends on who is making the evaluation, we offer relevant insights to debates about whether GenAI usage is acceptable. For example, our research suggests that students and teachers will have different perceptions when evaluating students' uses of GenAI to help them complete an assignment. Being aware of such self–other differences in inferred contributions is essential for an active, meaningful discourse about whether, when, and how GenAI usage is acceptable.

Second, this work contributes to research pertaining to perceptions surrounding intellectual property in technology realms (Jago & Carroll, 2023; Longoni et al., 2023), by exploring the extent to which people receive credit for an output if they obtain help from (Gen)AI versus humans. Unlike prior research, we do not seek to differentiate whether help came from (Gen)AI or humans; instead, we explore how people actually evaluate the tasks completed with the help of GenAI when they are considering their own versus others' work.

Third, by establishing this self–other difference in inferred usage of GenAI, we contribute to research dealing with self–other differences in general. This rich literature stream has documented such differences in various domains, such as evaluations of displays of vulnerability (Bruk et al., 2018) or smartphone usage behaviors (Barrick et al., 2022). We offer initial evidence of a self–other difference in inferred usage behavior of GenAI, with relevant consequences for the acceptability of such behavior.

## 2. Theoretical background

### 2.1. Research on GenAI

When it was introduced, society quickly recognized the importance of ChatGPT, as well as the new challenges it raises. Educational institutions continue to debate what kinds of rules to impose on ChatGPT usage (Fütterer et al., 2023), and some scientific journals have formulated guidelines for (in)appropriate uses for authors (Bockting et al., 2023). Companies also have identified some potential risks and sought to establish coherent guidelines to ensure transparency and data privacy (Fütterer et al., 2023; Sun, 2023). Similarly, governments have indicated intentions to regulate GenAI usage, though effective legislation will not be finalized for several years (Hutson, 2023). Thus, researchers are in a unique position, in that exploring research questions related to GenAI can help deepen understanding of both the technical features of such tools and the psychological and societal consequences of their usage (Barros et al., 2023; Bockting et al., 2023; Peres et al., 2023).

Early scientific investigations of GenAI usage and consequences have predominantly focused on ChatGPT. This research can be categorized broadly into two streams. First, research that explores the *objective performance* of ChatGPT endeavors to understand its capabilities and weaknesses. For example, Noy and Zhang (2023) demonstrate that ChatGPT can enhance writing productivity, offering a 40% reduction in average task completion time and an 18% increase in output quality when the tasks involve persuasive, generic writing. Other research identifies ChatGPT capabilities for learning "out-of-the-box thinking" and solving verbal insight problems, with performance levels equivalent to humans' (Orrù et al., 2023). It also out-



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

performs humans in emotional awareness (Elyoseph et al., 2023) and offers more empathetic advice than professional advice columnists (Howe et al., 2023). Still, some research in this stream emphasizes the shortcomings of ChatGPT, such as its potential to produce factually incorrect responses or fabricate bibliographic citations (Van Dis et al., 2023; Walters & Wilder, 2023).

Second, another research priority entails people's *subjective perceptions* of GenAI usage and its consequences, though this topic has received significantly less scientific attention thus far. As we have argued, a salient question in this realm is the extent to which people perceive intellectual property over output co-produced with the help of GenAI (Peres et al., 2023, Table 1, last column). The introduction of DALL-E prompted nearly immediate questions of "who owns the images" generated with the assistance of this tool (Goldman, 2023), and when students rely on ChatGPT to help with their assignments, instructors must determine the extent to which they can receive credit for the output. We know of no research that speaks directly to these questions, though Jago and Carroll (2023) find that creators (e.g., painters) receive more credit if they have been assisted by algorithms rather than other humans, because people infer that collaborating with algorithms requires more oversight by the creator than does collaborating with humans. Jago and Carroll (2023) take a third-person perspective to assess how others assign credit for output co-created with algorithms, whereas Longoni et al. (2023) focus on people's own perceptions of their ownership of AI-generated content. They find that people perceive plagiarized material generated by AI as less unethical than similar material generated by humans, because they attribute less ownership of the content to the AI. Similarly, Noy and Zhang (2023) determine that 80% of participants who outsourced writing tasks to ChatGPT simply copy-pasted its output. These findings suggest that people rely on ChatGPT (or other GenAI) when completing tasks but still feel ownership over the output.

### *2.2. Self–other differences in inferred contribution*

In addition to determining the amount of credit they should receive for their own output co-produced with GenAI, people also might need to evaluate others' work. According to literature on self–other differences, systematic differences arise in how people evaluate themselves versus others (e.g., Pronin, 2008). A classic illustration is the fundamental attribution error (Ross, 1977), that is, the tendency of observers to explain unfavorable behaviors of others (e.g., being late for a meeting)

**Table 1**
Overview of Studies.

| Study | Sample | Context | Main Findings |
|---|---|---|---|
| 1 | $N$ = 1158<br>U.S. representative sample | Writing, brainstorming and creativity, and learning and education | • Inferred contribution of the self (vs. others) is higher in all three domains (paradigm: imagined usage behavior). |
| 2a | $N$ = 363<br>U.S. convenience sample | Job application | • Inferred contribution of the self (vs. others) is higher (paradigm: real usage behavior). |
| 2b | $N$ = 928<br>U.S. convenience sample | Job application | • Inferred contribution of the self (vs. others) is higher regardless of whether evaluators experience the same task prior to evaluating another participant's work (paradigm: real usage behavior). |
| 3 | $N$ = 250<br>Dutch student sample | Inspiration and outsourcing behaviors | • Intentions to engage in any GenAI usage behavior are lower for the self (vs. others); the self–other difference in usage intention increases with the extent to which a behavior is considered outsourcing (paradigm: imagined usage behavior). |
| 4 | $N$ = 237<br>U.K. convenience sample | General usage | • The effect of the evaluation target on inferred contribution is mediated by perceptions of usage behavior (paradigm: imagined usage behavior). |
| 5a | $N$ = 251<br>U.K. convenience sample | Job application | • GenAI usage is considered more acceptable for the self (vs. others). The effect of the evaluation target on acceptability is mediated by inferred contribution (paradigm: imagined usage behavior). |
| 5b | $N$ = 601<br>U.S. student sample | Course assignment | • GenAI usage is considered more acceptable for the self (vs. others). The effect of the evaluation target on acceptability is mediated by inferred contribution (paradigm: imagined usage behavior). |
| A1 | $N$ = 244<br>U.K. convenience sample | General usage | • Inferred contribution of ChatGPT to a task is lower for the self (vs. others; paradigm: imagined usage behavior). |
| A2 | $N$ = 466<br>U.S. & U.K. teachers and students | Course assignment | • Students infer their own contribution to output co-produced with GenAI as higher compared with the credit they receive from teachers.<br>GenAI usage for student assignments is considered more acceptable by students than by teachers (paradigm: imagined usage behavior). |

Note: For Studies 2a and 2b, we report the $N$ on participant-level; note that our analyses are at the paragraph-level.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

according to dispositional and personality factors (e.g., the late attendee is a bad planner), but attribute their own unfavorable behavior to situational and environmental factors (e.g., the late attendee got stuck in traffic). Similarly, people think that products (e.g., medicine, online class, energy drink) more strongly influence the performance of others, compared with their own performance (Polman et al., 2022; Williams et al., 2012; Williams & Steffel, 2014). This effect extends to technology domains, such that when interaction partners use smartphones, it appears detrimental to a sense of social connection, but the person's own smartphone usage behavior does not evoke such perceptions (Barrick et al., 2022).

The various potential drivers of such self–other differences all appear rooted in two fundamental psychological processes: cognitive or motivational accounts. First, self–other differences may reflect cognitive differences in perceptions, which lead people to take fundamentally different information into account when evaluating themselves versus others (Chambers & Windschitl, 2004; Jones & Nisbett, 1972; Pronin, 2008; Watson, 1982). That is, we are aware of our own feelings and intentions, which precede, accompany, and follow our actions, but for others, we can only observe their actions. Therefore, we evaluate others on the sole basis of their observable behavior, but we evaluate ourselves also on the basis of what we think and feel (Pronin, 2008). Second, a motivation to maintain or enhance a positive self-view (Alicke, 1985; Brown, 1986, 2012; Kunda, 1990) leaves people more inclined to focus on their own positive intentions and attribute negative outcomes to contextual factors. Such self-enhancement motivations do not inform evaluations of others though.

In relation to our study context, we predict that people incorporate their own positive intentions into their evaluations of the level of their contribution to output co-produced with GenAI, but they do not account for positive intentions when evaluating others' contributions. Instead, people focus exclusively on observable behaviors, such as co-producing some outcome with a tool that could have done most of the work. These distinctive cognitive and motivational processes then lead people to perceive their own inferred contributions as systematically higher than the inferred contributions of others, a phenomenon we term "self–other bias in inferred contribution."

H1: People perceive their own contribution to an output they co-produced with GenAI as higher than others' contribution to an output that the others co-produced with GenAI.

Using GenAI to complete a task is typically not a binary decision but rather can vary in degree, so it allows for different inferences about the extent to which it has been used. This realization is essential to our prediction of a self–other bias in inferred contribution. For example, GenAI can be used to copy-edit self-generated text or to produce a rough first draft of a text; it also might be employed to fine-tune a self-generated research question or to produce the initial research question for someone. In these cases, the end product is the result of a collaboration between a human and GenAI, but the extent to which either party has contributed is not visible and may thus be subject to self–other perceptual differences.

As elucidated, to evaluate their own actions, people account for not just the observable outcome but also their own feelings and intentions, which preceded, accompanied, and followed the actions, often in an attempt to maintain a positive self-image (Kruger & Gilovich, 2004). Therefore, through both cognitive and motivational processes, we expect that people develop salient perceptions of their own intentions to complete tasks with no or limited help from GenAI, and they might regard its use as reflective of their intention to be inspired, but not solely to outsource the work. In contrast, when evaluating others' work, such intentions are unknown, so evaluators make assessments of the work based solely on the observable behavior, namely, the very production of the output in collaboration with a tool that could have done most of the work. In turn, the self–other bias pertaining to usage intentions should be weaker when the behaviors imply using GenAI for inspiration rather than for outsourcing the work to GenAI.

H2: People generally believe they are less inclined to use GenAI than others, and this difference in inferred usage intentions is more pronounced for behaviors considered outsourcing than for behaviors considered inspiration attempts.

This difference in the way that people regard the different uses of GenAI also might relate closely to the extent to which they give themselves, versus others, credit for output co-produced with GenAI. Therefore, we expect that differences in inferred usage behavior (ranging from inspiration to outsourcing) drive the self–other bias in inferred contribution.

H3: Differences in perceptions of usage behavior (on a continuum from inspiration to outsourcing) drive the self–other bias in inferred contribution, such that others' inferred contribution is lower because their usage behavior is perceived as more associated with outsourcing than with getting inspired.

Finally, the existence of such a self–other bias suggests that perceptions of the intellectual property over output co-produced with the help of GenAI are not universal but rather differ, depending on whether the evaluator created the output or is observing output created by others. The extent to which people infer that work has been created with the help of GenAI in turn should determine the extent to which they consider GenAI usage acceptable. We expect that GenAI usage appears more acceptable if the human contribution still is perceived as substantial, whereas it becomes less acceptable when the human contribution is perceived as relatively minor.

H4: Higher inferred contribution of the self (vs. others) to a task co-produced with GenAI leads to higher acceptability evaluations of GenAI usage.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

## 3. Overview of studies

Because ChatGPT has been central in many discussions surrounding the acceptability of GenAI tools, we focus on its usage in our empirical package. We test the hypotheses across nine studies (seven preregistered, total $N = 4498$; see Table 1), which document a consistent self–other bias in inferred contribution.

In detail, in Study 1, we document a self–other bias in three usage domains (writing, brainstorming and creativity, learning and education), using a nationally representative sample (H1). We replicate the effect with actual ChatGPT usage in Studies 2a and 2b, which further reveal that people infer a higher contribution of the self compared with others to a task involving the use of ChatGPT to write a paragraph for a job application. In Study 3, we identify greater self–other differences in intentions to use ChatGPT for outsourcing than for getting inspired (H2). We follow up on these results in Study 4 by documenting that differences in perceived usage behavior (i.e., inspiration versus outsourcing) drive the self–other bias in inferred contribution (H3). Studies 5a and 5b show that this bias in inferred contribution affects acceptability assessments (H4). Finally, we report two additional studies in the discussions of Studies 4 and 5.

All conditions and exclusions are reported; sample sizes were determined a priori (with details in the preregistrations). The data, R-code, materials, preregistrations, and IRB approvals are available at https://osf.io/qspwd/?view_only=3854679b-d33c4699bf4dab4fcb22c4cb. In addition, we list all the measures included for exploratory purposes in Web Appendix A. In line with the journal's requirements, we explicate our GenAI usage at the end of the text and also follow the living guidelines for transparency of GenAI usage formulated by Bockting et al. (2023), which we include at the end of this document.

## 4. Study 1

With Study 1, we conduct a first test of the self–other bias in inferred contribution across various domains, using a nationally representative sample and scenarios involving the imagined usage of ChatGPT.

### 4.1. Method

#### 4.1.1. Participants

A nationally representative sample of 1201 U.S. participants ($M_{age}$ = 45.69 years, 51% female) matched to the U.S. population distribution by age, gender, and ethnicity, was recruited from Prolific. As preregistered, participants ($N = 43$) who failed to respond correctly to an attention check question were excluded from the analysis, which left 1158 participants ($M_{age}$ = 45.94 years, 51% female).

#### 4.1.2. Design and procedure

Participants were randomly assigned to one of six conditions in a 2 (evaluation target: self vs. others) × 3 (usage domains: writing vs. brainstorming and creativity vs. learning and education) between-subjects design. First, we introduced ChatGPT to participants as an AI-powered chatbot that can aid with a variety of tasks. Participants then read about its capabilities in one of the three usage domains: writing, brainstorming and creativity, or learning and education. The descriptions of these capabilities were generated by ChatGPT, and for each domain, we purposefully included various example behaviors, spanning a continuum from getting inspired to outsourcing. Participants also learned that users have the flexibility to decide how much support they seek from ChatGPT. Next, participants in the self condition indicated how much they (vs. ChatGPT) would contribute to the task, if they were to use ChatGPT, whereas participants in the other condition indicated how much others (vs. ChatGPT) would contribute to the task if they were to use ChatGPT. Participants assigned 100 percentage points to ChatGPT and the evaluation target (self vs. others) as our measure of inferred contribution. After completing the attention check (see Web Appendix B), participants reported their age, gender, and ethnicity.

### 4.2. Results

A 2 × 3 analysis of variance (ANOVA) yielded a significant main effect for the evaluation target only ($F(1, 1152) = 122.58$, $p < .001$, $\eta_p^2 = .10$), such that the inferred contribution of the self is higher ($M_{self}$ = 54.62, SD = 27.64) than the inferred contribution of others ($M_{others}$ = 38.08, SD = 23.02). We did not find any main effect of the usage domain manipulation ($p = .14$) nor an interaction effect ($p = .28$). Planned contrasts indicate higher inferred contributions of the self versus of others in all three domains: writing ($M_{self}$ = 54.08, SD = 29.22; $M_{others}$ = 35.17, SD = 22.09, $t(1152) = 7.35$, $p < .001$, Cohen's $d = 0.74$), brainstorming and creativity ($M_{self}$ = 54.71, SD = 26.84; $M_{others}$ = 41.42, SD = 24.21, $t(1152) = 5.20$, $p < .001$, $d = 0.52$), and learning and education ($M_{self}$ = 55.00, SD = 27.07; $M_{others}$ = 37.58, SD = 22.22, $t(1152) = 6.52$, $p < .001$, $d = 0.70$; see Fig. 1).

### 4.3. Discussion

In line with H1, we detect a 17% higher average inferred contribution of the self versus of others. Thus, we gain initial evidence of a robust, generalizable self–other bias across a variety of usage behaviors and with a nationally representative sample.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

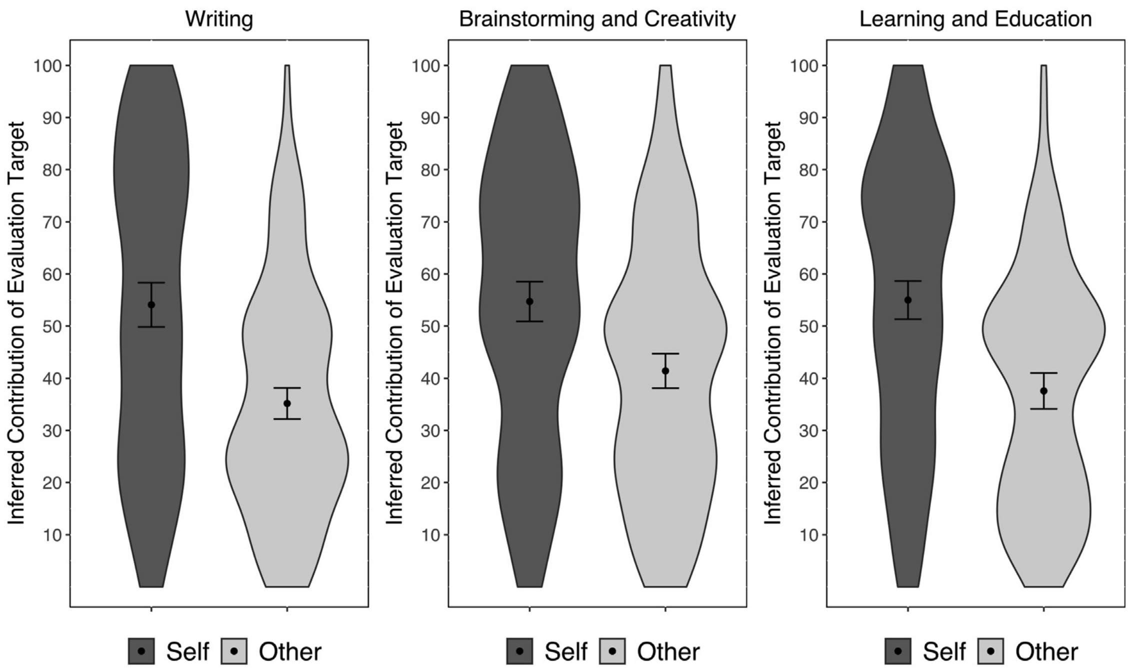


**Fig. 1. Inferred contribution of evaluation target (self vs. other) in three usage domains.** Notes: The violin plots represent the shape of the distribution of inferred contribution by experimental condition. The dot represents the mean and the error bars represent the 95 % confidence intervals.

## 5. Study 2

With Study 2, we fulfil two goals. First, we test the impact of the evaluation target on the inferred contribution, using a paradigm that allows for actual ChatGPT usage, utilizing a yoked design (Botti & Iyengar, 2004; Cordova & Lepper, 1996; Iyengar & Lepper, 1999; Klesse et al., 2019). Half of the participants (self condition) had to produce a paragraph about their own strengths and weaknesses for a job application, then indicated their contribution to the paragraph. The other half of the participants (other condition) read one such paragraph and inferred the contribution of the creator (i.e., another participant) to this paragraph. We predict that participants who wrote the paragraph would infer a higher contribution to themselves than participants who evaluated the exact same paragraph.

Second, the difference in inferred contributions arguably could be the consequence of different levels of experience with GenAI. Therefore, we replicate this study nine months after the initial data collection. In this replication, we added a third condition, in which we asked participants assigned to the other-experienced condition to write the application paragraph themselves, before evaluating the work of another participant. By comparing these two studies, we can test the role of gaining experience with the focal task as a creator, prior to taking on the role of an evaluator, and determine if any changes to the self–other bias arise over time.

### 5.1. Method

#### 5.1.1. Participants

In Study 2a, we assigned 201 U.S. participants ($M_{age}$ = 40.50 years, 49% female) to batch 1 and 200 U.S. participants ($M_{age}$ = 32.74 years, 42% female) to batch 2. In Study 2b, 199 participants ($M_{age}$ = 38.62 years, 51% women) were in batch 1, and 799 participants ($M_{age}$ = 39.50 years, 47% female) were in batch 2 (with 410 participants in the other condition and 389 participants in the other-experienced condition). All participants were recruited from Prolific.

#### 5.1.2. Design and procedure

The data were collected in two batches, in accordance with a yoked design. That is, participants produced a paragraph outlining their strengths and weaknesses, which represents the self condition. After all these data for batch 1 were collected, the paragraphs were automatically fed into a new Qualtrics survey. The batch 2 data were collected on the same or the next day. In batch 2, participants evaluated one unique paragraph, produced by a participant from batch 1, which represents the other condition. Study 2a includes a single other condition (i.e., evaluating the paragraph without completing the focal task), but in Study 2b, we had two other conditions: one without experience (replication of Study 2a) and one with prior experience with the focal task (other-experienced condition).



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006


**Fig. 2. Inferred contribution of evaluation target (self vs. other) to the paragraphs produced.** Notes: The violin plots represent the shape of the distribution of inferred contribution by experimental condition. The dot represents the mean and the error bars represent the 95% confidence intervals. Study 2a is presented in panel a; Study 2b is presented in panel b.

Participants in both batches learned that the survey consists of two parts that they would complete in one sitting. In the first part (same for batches 1 and 2), participants were introduced to ChatGPT, then instructed to open a ChatGPT account and write a poem with its help. Thus, we ensured that ChatGPT usage was possible (and salient) for all participants. After completing this first part, they moved to the second part, which represented the focal task for this study. In batch 1 (self condition), participants had to imagine they were applying for a job position in a prestigious firm and needed to write a short paragraph describing their strength and weaknesses. They wrote this short paragraph (minimum requirement of 1200 characters). Due to the poem writing task in part 1, ChatGPT usage was salient and accessible for all these participants, but we did not explicitly mention using ChatGPT in part 2. In batch 2 (other condition), participants learned that other people had to write a paragraph about their strengths and weaknesses and that it was their task to read and evaluate one such paragraph. Participants in the other-experienced condition (Study 2b) first completed the paragraph writing task themselves, then read and evaluated one paragraph written by another participant.

After writing/reading the application paragraph, all participants responded to the inferred contribution measure and evaluated how much they (self condition) or the other participant (other conditions) and ChatGPT contributed to creating the paragraph, by dividing 100 percentage points between ChatGPT and the human (self or other). Finally, they responded to an attention check (Web Appendix B) and reported their age and gender.[2]

### *5.2. Results*

#### *5.2.1. Data preparation*

Unlike Study 1, we conducted our analyses not at the level of the participant but rather at the level of the paragraph. Thus, we only retained paragraphs evaluated in both the self and the other conditions. In contrast, we excluded any paragraphs written in batch 1 if they were not matched with any participant in batch 2 (other conditions), as well as any observations that had to be excluded because participant(s) failed the attention check (Web Appendix C). If paragraphs were evaluated by more than one participant in the other conditions, we took the average. Accordingly, our analyses are based on 167 paragraphs for Study 2a and 181 paragraphs for Study 2b.

## 6. Inferred contribution

In Study 2a, the paired samples *t*-test (at the paragraph level) reveals a higher inferred contribution of the self ($M_{self}$ = 55.84, SD = 39.54) compared with of the other ($M_{other}$ = 35.15, SD = 24.91, $t(166) = 6.34$, $p < .001$, $d = .63$) (Fig. 2, Panel a). In Study 2b, a repeated measures ANOVA, with the three conditions (self, other, and other-experienced) as predictors and inferred contribution as the outcome variable, yields a significant main effect of the evaluation target ($F(2, 360) = 32.12$, $p < .001$, $\eta_p^2 = .15$). The planned contrasts also reveal higher inferred contributions in the self condition ($M = 59.10$, SD = 39.28) than in the other condition ($M = 41.31$, SD = 21.90, $t(360) = 7.46$, $p < .001$, $d = 0.56$) and in the other-

[2] We also included some exploratory measures in our studies, such as participants' experience with ChatGPT (measured in most studies), perceived quality of the paragraphs produced (Studies 2a and 2b), perceptions of psychological ownership of the outcome produced (Studies 4, 5a, and A1), feelings of deservingness and pride, and chances to obtain similar outcome without the help of ChatGPT (Study 5a). The results of analyses that include these exploratory measures are detailed in Web Appendix A.

experience condition ($M$ = 44.14, SD = 25.39, $t(360) = 6.27$, $p < .001$, $d = 0.45$) conditions (Fig. 2, Panel b). There was no difference between the other conditions ($p = .236$).

#### 6.0.1. Exploratory analyses

Arguably, others might be incapable of discriminating different degrees of creators' contributions to the paragraphs, or they might offer random responses. To test these possibilities, we assess the correlation between the inferred contribution of the self and of others; it is significant (Study 2a $r(165) = .21$, $p = .008$; Study 2b $r_{self\text{-}other}(179) = .45$, $p < .001$; $r_{self\text{-}other\ experience}(171) = .51$, $p < .001$). That is, others appear capable of accurately detecting variation in a creator's contribution when using ChatGPT. Yet, the average mean difference of 18% continues to indicate a self–other bias in inferred contribution.

Due to the designs of Studies 2a and 2b, we can pool the data and include time as an additional factor in our analysis. In turn, a 2 × 2 mixed ANOVA with evaluation target (self vs. other) at the paragraph level and time (Study 2a vs. Study 2b; nine month time difference) as between-subjects variables reveals a significant main effect of the evaluation target ($F(1, 346) = 85.48$, $p < .001$, $\eta_p^2 = .19$) and a marginally significant main effect of time ($F(1, 346) = 2.86$, $p = .092$, $\eta_p^2 = .01$), but no significant interaction between them ($p = .487$).

### 6.1. Discussion

With Study 2, we offer two contributions. First, we establish the external validity of the self–other bias by generalizing it to real usage behavior. The results offer further support for H1, such that the inferred contribution of a person to a certain output (e.g., application letter) is perceived as higher when people evaluate their own work compared with when they evaluate the work of others. On average, we detect a mean difference of 18%, comparable to the mean difference of 17% we detected in Study 1.

Second, we explore whether experience with GenAI affects the self–other bias in inferred contribution. When we replicated Study 2a nine months later, we identified a self–other bias similar in effect size (though slightly smaller); it did not change meaningfully even though ChatGPT adoption increased substantially. Moreover, the inferred contribution to others does not seem influenced by experience with the task, considering that it is comparable in magnitude irrespective of whether evaluators completed the focal task themselves prior to evaluating the work of creators. Notably, the correlation between the inferred contributions of the self and of others even increased over time (from $r = .21$ to $r = .45$; $z = -2.55$, $p = .010$), implying that over time, evaluators may have grown more accurate in detecting variation in the extent to which ChatGPT has been used.

## 7. Study 3

The aim of Study 3 is to provide evidence related to our prediction that people infer different types of usage behavior for themselves versus others. As outlined in H2, we expect that the difference in inferred contribution arises because people think that they use GenAI only for inspiration, whereas others use it for outsourcing tasks. Therefore, with this study, we measure participants' own intentions or their perceptions of others' intentions to engage in behaviors that vary in the extent to which they imply using GenAI for outsourcing versus for inspiration. If our expectations hold, we should find consistently lower intentions to engage in any GenAI usage behavior for the self than for others, as well as a bigger difference between the self and others for tasks considered outsourcing than for tasks considered getting inspired.

### 7.1. Method

All data ($N$ = 346) were collected in one batch in a university lab. We dedicated day 1 to the collection of pretest data ($N$ = 83, $M_{age}$ = 19.34 years, 52% female), and the rest of the days to the collection of data for our main study ($N$ = 263, $M_{age}$ = 19.15 years, 47% female). All analyses were conducted after the complete data collection was finished. Thirteen participants who only partially completed the survey were excluded from the analysis.

#### 7.1.1. Pretest

Participants categorized 10 behaviors as indicative of using ChatGPT to get inspired or using ChatGPT to outsource the task. We used ChatGPT to generate these 10 behaviors (e.g., "Have ChatGPT provide solutions to a homework assignment"). The pretest results reveal the proportion of participants who categorize each behavior as outsourcing or getting inspired, which we consider an indication of the extent to which each behavior is considered outsourcing a task to ChatGPT. In Web Appendix D (Table D.1), we list all the tasks, together with the percentage of participants categorizing each behavior as outsourcing.

#### 7.1.2. Main study

The study employed a mixed design, with evaluation target (self vs. others) as a between-subjects factor and the extent to which each behavior is considered outsourcing as a continuous, within-subjects variable (score obtained in the pretest). Half of the participants had to indicate their own likelihood of engaging in each of the 10 behaviors (self condition), and the other



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

half inferred the likelihood that another student at their institution (other condition) would engage in these behaviors (7-point scale; 1 = "not at all likely," 7 = "very likely"). The participants also reported demographic information.

### 7.2. Results

#### 7.2.1. Likelihood to engage in the behavior

We ran a mixed-effect regression model, with evaluation target and the extent to which each behavior is considered outsourcing (percentage score obtained in pretest, mean centered) and their interaction as predictors. Likelihood to engage in each behavior is the dependent variable. We also account for the within-subject rating of behaviors using a participant-specific random intercept. The main effect of evaluation target ($b$ = −1.04, SE $b$ = .12, $t$(2498) = −8.66, $p$ < .001), the main effect of the extent to which each behavior is considered outsourcing ($b$ = −.01, SE $b$ = .001, $t$(2498) = −7.15, $p$ < .001), and their interaction ($b$ = −.008, SE $b$ = .002, $t$(2498) = −3.74, $p$ < .001) all are significant. A floodlight analysis affirms that this difference is significant at all values, regardless of the extent to which each behavior is considered outsourcing. However, the significant interaction effect suggests that the self–other bias in inferred usage likelihood grows larger the more a behavior is considered outsourcing (Fig. 3).

### 7.3. Discussion

In line with H2, Study 3 shows that people think that others are relatively more likely to use GenAI for any task, as well as that they predominantly do so for tasks that imply outsourcing to GenAI. Therefore, this study establishes evidence of the proposed underlying process.

## 8. Study 4

Study 4 seeks to offer more direct evidence of the proposed underlying process for the self–other bias in inferred contribution that we document in Studies 1 and 2. To this end, we measure whether differential inferences of the types of usage behavior (ranging from inspiration to outsourcing) mediate the self–other bias in inferred contribution.

### 8.1. Method

#### 8.1.1. Participants

We recruited 252 U.K. participants ($M_{age}$ = 40.46 years, 50% female) from Prolific. As preregistered, participants ($N$ = 15) who failed the attention check were excluded from the analysis, leaving 237 participants ($M_{age}$ = 40.26 years, 48% female).

#### 8.1.2. Design and procedure

Participants were randomly assigned to the self condition or other condition in a between-subjects design, then presented with general information about ChatGPT and its capabilities. The participants had to estimate how much they (vs. another person) would have contributed to a task (no further specifications) if they were to use ChatGPT to complete it. As in Studies

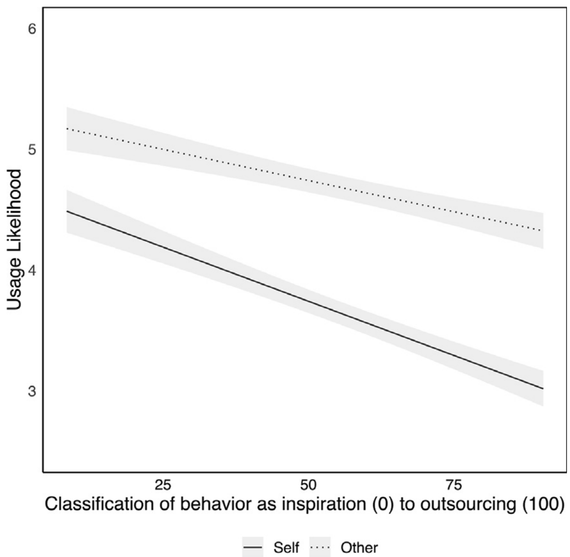


**Fig. 3.** Assessments of the likelihood to use ChatGPT for outsourcing.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

1 and 2, participants responded by distributing 100 percentage points between ChatGPT and the evaluation target (self vs. others) to indicate the contributions. Next, participants indicated their agreement with three items that we used to measure perceptions of usage behavior: "When using ChatGPT, I [others] only use it to get inspired/do not directly copy anything ChatGPT suggests/still put in my [their] own work" (7-point scales, 1 = "strongly disagree," 7 = "strongly agree"; Cronbach's $\alpha = .72$). Higher scores on this composite measure suggest that ChatGPT is used more for inspiration rather than outsourcing the task. All participants then completed the attention check and reported their age and gender.

### 8.2. Results

An independent samples *t*-test confirms higher inferred contributions of the self ($M_{self} = 53.36$, SD = 28.95) than of others ($M_{others} = 31.01$, SD = 22.34, $t(235) = 6.65$, $p < .001$, $d = 0.86$). Then another independent samples *t*-test shows that participants perceive themselves as more likely to utilize ChatGPT for inspiration and help, whereas they regard others as more likely to outsource the task ($M_{self} = 4.84$, SD = 1.20; $M_{others} = 3.49$, SD = 1.07, $t(235) = 9.15$, $p < .001$, $d = 1.19$). In line with our predictions, Process Model 4 (Hayes, 2013; 5000 bootstrap samples) confirms that the effect of the evaluation target on inferred contribution is mediated by perceptions of usage behavior ($b = 16.62$, 95 % CI [12.07, 21.77]). The differential perceptions of usage behavior thus explain why participants infer a higher contribution of the self (vs. others). See Fig. 4.

### 8.3. Discussion

We rooted our expectation of higher inferred contribution of the self versus others in prior literature on self–other differences, which documents that when people evaluate others, they focus on observable behaviors, whereas self-evaluations also take intentions and feelings into account (Pronin et al., 2002). In the context of GenAI usage, this variation implies that we evaluate others predominantly on the observable behavior of using the tool, which can perform the task, whereas evaluations of ourselves also account for (good) intentions to seek inspiration. Study 4 confirms this expectation and shows that differential inferences of usage behavior drive the self–other bias in inferred contribution.

One question that may arise is whether differences in inferred contribution mainly lie with ChatGPT or with the evaluation target, which was confounded in our previous studies where we only used one measure (i.e., dividing 100 points between the evaluation target and ChatGPT). To address this concern, we replicated Study 4 with two separate measures of inferred contribution (one for ChatGPT and one for evaluation target; see Web Appendix E). It reveals effects of the evaluation target on both measures; that is, participants infer a higher contribution of the self than of others, but they also infer a lower contribution of ChatGPT to the task when used by the self versus others. These results suggest that people consider the contribution of both parties (person and ChatGPT) to a task not independently but in a complementary way. Furthermore, this study replicates the effect on types of usage behavior (i.e., inspiration versus outsourcing), which mediates the effect of the evaluation target on inferred contribution.

## 9. Study 5

With Study 5, we test whether the self–other bias in inferred contribution has downstream consequences for the extent to which people consider ChatGPT usage acceptable. We test this prediction in two contexts: writing a letter for a job application (Study 5a) and completing a student assignment (Study 5b).

### 9.1. Method

#### 9.1.1. Participants

For Study 5a, we recruited 251 U.K. participants ($M_{age} = 39.45$ years, 51% female) from Prolific. All participants responded to the attention check question correctly and remained in the analysis. For Study 5b, we recruited 603 U.S. students ($M_{age}$ =

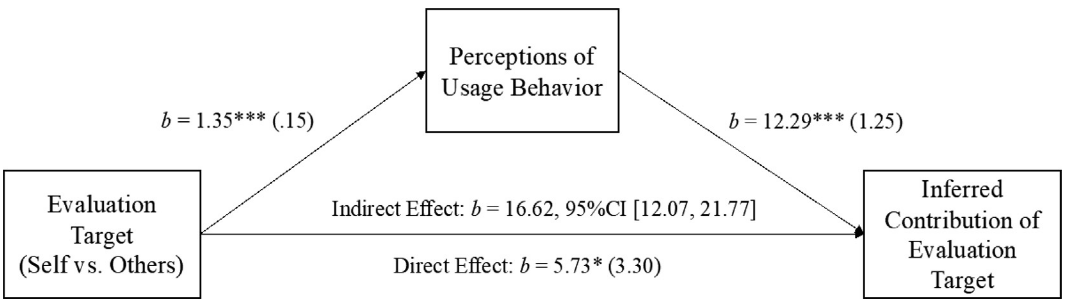


**Fig. 4. Effect of evaluation target on inferred contribution, mediated by perceptions of usage behavior.** Notes: The evaluation target is coded 1 for the self and 0 for others. Perceptions of usage behavior range from 1 (=outsourcing) to 7 (=inspiration). Inferred contribution of the evaluation target ranges from 0% to 100%. *$p < .05$, ** $p < .01$, *** $p < .001$.

28.14 years, 47% female) from Prolific. Participants ($N$ = 2) who failed to respond to the attention check correctly were excluded, leaving 601 participants ($M_{age}$ = 28.07 years, 47% female). Participants also confirmed that they were students and noted where they studied.

### 9.1.2. Design and procedure

Participants were randomly assigned to either the self condition or the other condition in a between-subjects design in both studies. Those participants assigned to the self conditions imagined they were applying for a job in a highly competitive industry, used ChatGPT to prepare an application letter, and earned the position (Study 5a), or that they used ChatGPT to complete a final assignment for a competitive course in which the professor grades on a curve, and they were among the top 5% of students (Study 5b). Participants in the other conditions imagined another person applying for a job in a highly competitive industry, using ChatGPT to prepare the letter, and obtaining the position (Study 5a) or another student having used ChatGPT to complete a final assignment for a competitive course and being among the top 5 % of students (Study 5b).

All participants then responded to the inferred contribution measure (as in Studies 1, 2, and 4). Next, they rated how acceptable it was for the evaluation target (self vs. other) to use ChatGPT for the task on a 7-point scale (1 = "not at all," 7 = "very much"). They also rated their agreement with the perceived usage behavior items from Study 4 (Cronbach's $\alpha$ = .84 for both studies), completed the attention check, and reported their age and gender.

## 9.2. Results

Replicating our previous findings, two independent samples $t$-tests show that the inferred contribution of the self is higher than the inferred contribution of the other person, in both Study 5a ($M_{self}$ = 52.33, SD = 27.86; $M_{others}$ = 37.34, SD = 24.92, $t(249)$ = 4.50, $p$ < .001, $d$ = 0.57) and Study 5b ($M_{self}$ = 54.55, SD = 31.46; $M_{others}$ = 40.71, SD = 28.23, $t(599)$ = 5.68, $p$ < .001, $d$ = 0.46). Then another set of independent samples $t$-tests affirms that participants perceive themselves as more likely to utilize ChatGPT for inspiration rather than for outsourcing compared with others, again in both Study 5a ($M_{self}$ = 4.68, SD = 1.49; $M_{others}$ = 3.72, SD = 1.48, $t(249)$ = 5.16, $p$ < .001, $d$ = 0.65) and Study 5b ($M_{self}$ = 4.71, SD = 1.76; $M_{others}$ = 3.64, SD = 1.56, $t(599)$ = 7.85, $p$ < .001, $d$ = 0.64). The mediation analyses further confirm that usage behavior drives the effect of the evaluation target on the inferred contributions in both studies (see Web Appendix D).

Importantly, another independent-samples $t$-test shows that participants in Study 5a view ChatGPT usage as more acceptable for themselves ($M_{self}$ = 4.28, SD = 1.76) than for others ($M_{others}$ = 3.80, SD = 1.88, $t(249)$ = 2.09, $p$ = .038, $d$ = 0.26; Fig. 5). Process Model 4 (Hayes, 2013; 5000 bootstrap samples) establishes that this effect of the evaluation target on acceptability is mediated by inferred contribution ($b$ = .35, 95 % CI [.17,.55]). Study 5b replicates the main effect of the evaluation target on acceptability ($M_{self}$ = 3.57, SD = 2.14; $M_{others}$ = 3.21, SD = 1.91, $t(599)$ = 2.14, $p$ = .033, $d$ = 0.17), and Process Model 4 (Hayes, 2013; 5000 bootstrap samples) again confirms the mediation by inferred contribution ($b$ = .10, 95 % CI [.02,.19]). Web Appendix D (Fig. D.2) contains the mediation models for Studies 5a and 5b.

## 9.3. Discussion

We find that the self–other bias in inferred contribution has ramifications for the extent to which people consider GenAI usage acceptable, leading to higher acceptability evaluations for the self (vs. others). These findings are particularly insightful

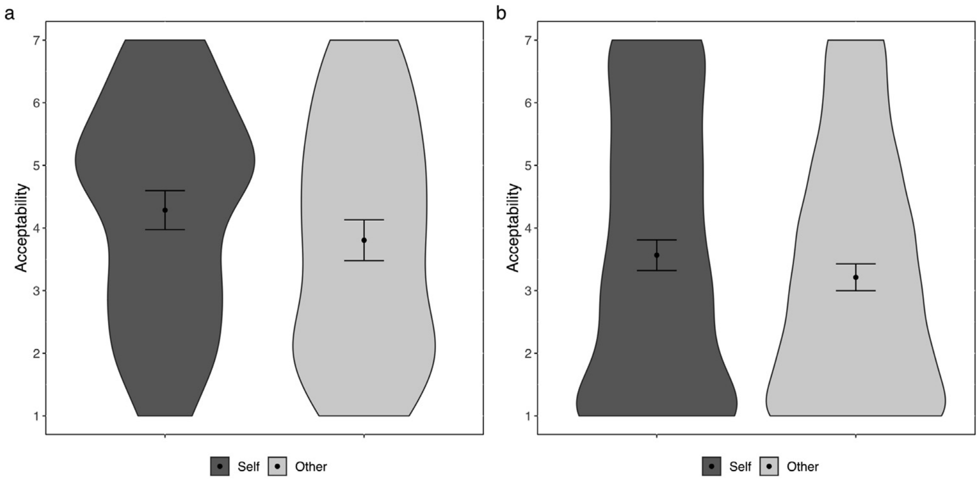


**Fig. 5. ChatGPT usage appears more acceptable for the self than for others.** Notes: The violin plots represent the shape of the distribution of acceptability by experimental condition. The dot represents the mean, and the error bars represent the 95% confidence intervals. Study 5a is presented in Panel a; Study 5b is presented in Panel b.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

for ongoing debates at educational institutions about whether GenAI usage is acceptable for students or not. First, they suggest that the divergent stances taken by students and teachers may reflect their distinct perceptions of the extent to which a student contributes and thus should receive credit for output co-created with GenAI. Second, they imply that acceptability of GenAI usage is higher if it is predominantly utilized for behaviors associated with using GenAI for inspiration (rather than outsourcing).

Admittedly, we studied students evaluating either their own usage behavior or the usage behavior of their peers, while in practice the discussion typically involves different perspectives of students versus teachers. Whereas our study paradigm allowed for random assignment (to the self vs. other condition) which fosters internal reliability, it might have come at the cost of external reliability. Therefore, we ran a conceptual replication of Study 5b, for which we recruited participants from two separate samples: students and teachers. Students evaluated their own contribution to an assignment that could have been completed with ChatGPT, whereas teachers evaluated the contribution of students to this assignment. The results confirm the existence of self–other biases in both inferred contribution ($M_{student}$ = 68.68, SD = 28.88; $M_{teacher}$ = 59.28, SD = 24.72, $t(464)$ = 3.76, $p$ < .001, $d$ = 0.35) and the acceptability of ChatGPT usage ($M_{student}$ = 3.46, SD = 2.04; $M_{teacher}$ = 2.48, SD = 1.52, $t(464)$ = 5.80, $p$ < .001, $d$ = 0.54), suggesting the ecological validity of our findings (see Web Appendix E).

## 10. General discussion

The emergence and widespread adoption of GenAI tools already has had (and will continue to have) profound impacts on society, business practices, education, and science. The rapid diffusion of ChatGPT in particular has prompted (heated) discussions about the acceptability of its use to complete tasks (Fütterer et al., 2023; Peres et al., 2023; Thorp, 2023). With this research, we contribute to this discussion by documenting that people assign more credit to themselves than to others for output co-produced with GenAI, a phenomenon that we term "self–other bias in inferred contribution." Across nine experiments ($N$ = 4498; Web Appendix F summarizes the results), we offer critical answers to the following questions that are pivotal to ongoing debates surrounding the extent to which people (should) use GenAI.

*How generalizable is this self–other bias in inferred contribution?* The results of all our experiments point to a large average effect size ($d$ = 0.55) across different populations (various nationalities, recruited from panels and a student pool) and experimental paradigms. We observe the effect with both actual and imagined ChatGPT usage and across different usage domains. Furthermore, the effect is robust to different constellations of the self–other construct, such as students' ChatGPT usage being evaluated by fellow students or teachers. Thus, the self–other bias appears to persist even if evaluators are in the same position (i.e., two students) and when evaluating others' work is part of their daily role (i.e., teachers and students).

A descriptive assessment also reveals that people infer that ChatGPT contributes significantly to a task, even if they make the judgment for themselves. Overall, the mean inferred contributions center consistently at around 54% for the self and 38% for others. It may seem shocking that people sense that they outsource (more than) half of the task to ChatGPT; this finding implies that people are quickly becoming accustomed to letting ChatGPT do a substantial proportion of their work.

*How does this self–other bias relate to debates about the acceptability of GenAI usage?* The debate about the acceptability of GenAI includes both advocates and critics, each side with (valid) arguments for why GenAI usage should (not) be acceptable. With this research, we cannot offer conclusions regarding who is right or wrong, but it does specify that the evaluation target drives people's acceptability evaluations. In this sense, we posit that people's stances on acceptability are not black and white but rather that they may be prone to applying double standards, being harsher on others than on themselves.

*Can evaluators accurately detect GenAI usage?* When discussing the difference between creators' and evaluators' assessments of GenAI usage, it is important to understand whether evaluators can detect various levels of GenAI usage accurately. To shed light on this question, we consider two metrics. First, in Studies 2a and 2b, we find a strong correlation between the relative contribution evaluations in the self and other conditions, which contradicts the possibility of random patterns exhibited by evaluators. Second, we can obtain an independent estimate of the extent to which ChatGPT has contributed to output by using widely available AI detector or checker tools. Even if such tools are not 100% accurate, we can obtain estimates of the relative contributions of the human versus ChatGPT to each of the paragraphs written for Study 2. We used GPTZero (https://gptzero.me/) which gives us estimates of the relative contribution of the human versus ChatGPT to each paragraph.[3] The judgment of the AI detector tool of the human contribution was significantly and highly correlated with inferred contributions in the self condition (Study 2a $r$ = .84; Study 2b $r$ = .81), suggesting the validity of its estimates. Further, we observed a significant, though weaker, correlation between the tool's estimate and the inferred contribution in the other conditions (Study 2a $r$ = .26; Study 2b $r_{other}$ = .52; $r_{other\ experienced}$ = .57), which signals the accuracy of the evaluators in detecting various levels of GenAI usage. We also note that evaluators seem to be growing more accurate over time in discriminating different levels of contributions to output co-created with GenAI, as evidenced by stronger correlations between the estimate in the other conditions and the tools' estimate in Study 2b versus Study 2a ($z_{2a\text{-}2b\ control}$ = -2.88, $p$ = .004; $z_{2a\text{-}2b\ other\ experience}$ = -3.51, $p$ < .001).

*Who is biased?* Identifying the source of the bias is important for establishing appropriate debiasing efforts. To gain initial insights, we undertake a further comparison of the results obtained by the tool with the estimates provided by our participants. For Study 2a, it reveals a positive difference between the creators' and the tool's estimate ($M_{self}$ = 55.84, SD = 39.54; $M_{tool}$ = 41.15, SD = 44.51; $t(166)$ = 7.93, $p$ < .001, $\Delta$ = 14.69) and a (marginally significant) negative difference between the

[3] To validate the estimates provided by the tool, we fed each paragraph into it twice, which produced small variations in the estimates but strong correlations between them. All analyses reported here are based on the average values.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

evaluators' and the tool's estimate ($M_{others}$ = 35.15, SD = 24.91; $t(166) = -1.72$, $p = .087$, $\Delta = -6.00$). However, for Study 2b, we only observe a positive difference between the creators' and the tool's estimate ($M_{self}$ = 59.10, SD = 39.28; $M_{tool}$ = 40.24, SD = 45.89; $t(180) = 9.40$, $p < .001$, $\Delta = 18.86$), with no difference between the evaluators' and the tool's estimate ($M_{other}$ = 41.31, SD = 21.90, $t(180) = .37$, $p = .715$, $\Delta = 1.07$; $M_{other\ experience}$ = 44.14, SD = 25.39, $t(180) = 1.39$, $p = .167$, $\Delta = 3.90$). These results suggest a bias on the side of the creator, who underestimates the GenAI's contribution. However, it is less clear whether bias exists on the evaluator side. Study 2a implies both sources of bias (even if of different magnitudes), but bias coming from the evaluator seems to have disappeared in Study 2b.

## 11. Theoretical contributions

The main theoretical contribution of this work lies in deepening our understanding of how people evaluate their own (vs. others') contributions to work co-produced with GenAI. Despite the hype surrounding GenAI and calls for scientific research on this matter (Davenport et al., 2020; Peres et al., 2023), empirical investigations are still sparce and focused mainly on its objective capabilities (e.g., Elyoseph et al., 2023; Howe et al., 2023; Noy & Zhang, 2023; Orrù et al., 2023). Some prior work also compares people's responses to humans with their responses to algorithms (Castelo et al., 2019, Longoni et al., 2019), GenAI (Bergner, et al., 2023; Zhang & Gosline, 2023; Zhang et al., 2024), or chatbots (Castelo et al., 2023). We advance such contributions in several ways.

First, we focus on subjective perceptions surrounding GenAI usage, in instances in which people co-produce output with GenAI, so we can establish whether usage perceptions depend on *who* interacts with it (the self versus others). In doing so, our work is the first to point to a fundamental self-other bias in the inferred contribution when co-producing output with GenAI, and its consequences for acceptability decisions of GenAI usage. Documenting this self–other bias provides a direct response to questions about perceptions of intellectual property raised by Peres et al. (2023).

Second, we add a different perspective to research on intellectual ownership that has mostly compared work generated with the help of either AI or other humans (Jago & Carroll, 2023; Longoni et al., 2023). With ChatGPT usage quickly becoming a normal practice, it is pivotal to shift our research focus to deepening our understanding of factors that shape human-AI interactions and influence people's perceptions surrounding it.

Third, we contribute to research on self–other differences by pointing to the evaluation target as a fundamental factor that determines perceptions of GenAI usage behavior and acceptability. In turn, this research aligns with recent work by Agarwal et al. (2024) documenting self–other differences in the adoption of automated vehicles. Together, the insights emphasize the relevance of studying self–other differences in relation to new technology adoption likelihood and usage behaviors.

### *11.1. Practical implications*

Considering how widespread GenAI usage is already, across all layers of society, and how disruptive it can be to education, business, and science, the practical implications of our work pertain to a wide range of stakeholders. Table 2 zooms in on implications for the stakeholders most relevant to the target audience of this journal. For each of these stakeholders we present the current status quo (i.e., common knowledge or prevalent practices), the most important insights from our research, and some clear action points for how to alter the current state of affairs, if applicable. In this section, we refrain from reiterating the points included in the table and instead offer some concrete examples and suggestions for several stakeholder groups. We also discuss a few salient themes pertinent to our research findings.

*Educators and students.* The acceptability debates sparked by the introduction of GenAI might seem surprising, considering that today's students always have had access to tools that help them complete assignments (e.g., Google, spelling checkers). The disruptive nature of GenAI for education seems to reflect its ability to complete tasks with limited contributions from creators. Our study suggests that it is exactly this type of usage behavior that is widely considered unacceptable. An immediate response might be to forbid GenAI usage. An alternative is to find ways to encourage students to use GenAI for inspiration but not outsourcing. For example, Khan Academy introduced an AI tool called Khanmigo that offers students inspiration and help but will not complete their tasks (Khan Academy, 2023).

*Marketing professionals.* Across business functions, GenAI usage is most common in marketing and sales (Chui et al., 2023), but within these domains, various subspecialities likely differ in the extent to which they can benefit from GenAI. For example, content marketers might leverage GenAI to write blog and social media posts or to generate headlines and drafts. Marketing research agencies can apply GenAI to analyze market trends. Influencer marketers may employ GenAI to identify potential influencers that fit their brand and targeting strategies. For all of these tasks, using GenAI promises significant time and cost savings (Acar, 2023). However, our research suggests that customers might underestimate the human contributions to the output if marketing agencies employ GenAI, which in turn might decrease their valuation of such services. In this sense, GenAI usage could threaten the perceived relevance of the marketing profession as a whole. But other subspecialities might evoke perceptions that they use GenAI mainly for inspiration instead of outsourcing, such as marketing strategists. In these cases, customers may infer higher contributions by humans, with less severe ramifications for the perceived relevance of this category of marketing jobs.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

**Table 2**
Implications of the current findings for various stakeholders.

| | Educators and students | Scientific journals and authors | Marketing professionals | Tech developers | Policy makers and regulators | General public |
|---|---|---|---|---|---|---|
| **Status quo** | • Debate on whether students may use GenAI for completing assignments, with opinions ranging from accepting to abandoning GenAI. | • Few journals have published guidelines and rules on how to use and disclose GenAI usage in scientific research, which differ in their level of specificity.<br>• The living guidelines (from *Nature*) refer to accountability (human contribution) and transparency (disclosure of GenAI usage) as key principles. | • GenAI usage can increase quality and efficiency of companies (e.g., developing new products, marketing campaigns), but disclosure of GenAI use can undermine the perceived value of the company. | • The focus is predominantly on the technical side of things (i.e., developing the best GenAI tools). | • Policy makers predominantly focus on the outcomes created by GenAI (quality and risks).<br>• Discussions on ethical and social implications (privacy, bias, and discrimination), transparency, and accountability focus on the outcome that GenAI can produce and the underlying algorithms that generate it. | • A broad population using GenAI, potentially being unsure whether and when GenAI usage is acceptable. |
| **Insights from the current research** | • Creators and evaluators of an outcome co-produced with GenAI start from different assumptions about contributions to the task.<br>• Creators underestimate GenAI's contribution, whereas evaluators may be more accurate or overestimate GenAI's contribution.<br>• The differential inferences regarding contributions stem from the assumption that others (self) use GenAI for outsourcing (inspiration). | • Creators may underestimate the extent to which they use GenAI.<br>• Creators may think that using GenAI for inspiration is acceptable and, therefore, may not require disclosure. | • Customers may question companies' added value because they infer high contribution of GenAI in delivering outcomes. | • The same objective outcome is construed differently in terms of GenAI's (vs. users') contribution to it by creators and evaluators.<br>• Using GenAI is not a yes/no decision but offers a wide range of usage possibilities that leave room for interpretation and biases on both the evaluator and creator sides. | • Policy makers should be aware that a mere focus on the outcomes that GenAI can produce, the underlying data, and algorithms is insufficient; considering the humans evaluating the outcome is also important.<br>• Policy makers should be aware that their stance in the debate may differ depending on their own role (i.e., creator or evaluator). | • Answer to the question of acceptability is not universal but depends on whether one decides for oneself or others.<br>• People's perceptions surrounding usage behavior and overall contributions to a task are biased, in that they assign more (less) credit to themselves (others). |
| **Action points** | • Make educators and students aware of this self–other bias in assigning credits for an outcome co-produced with GenAI.<br>• Educate students on different usage behaviors (outsourcing vs. inspiration) and their acceptability.<br>• Write clear guidelines on acceptability of GenAI usage that take into account the perspectives of creators and evaluators, as well as different types of usage behavior. | • Make editors, reviewers, and authors aware of this self–other bias in assigning credit for an outcome co-produced with GenAI.<br>• Broaden perceptions of the scope of GenAI usage behavior that requires transparency (i.e., all usage behavior should be disclosed).<br>• Specify disclosure guidelines that leave less room for interpretation by the creator about what needs to be disclosed and by the evaluator about how GenAI has been used. | • Companies can protect their sustainable competitive advantage by being transparent about their added value (beyond that of GenAI tools). | • GenAI interfaces could implement a feedback system that evaluates the relative contribution of GenAI (vs. the user) to the final outcome.<br>• Technology developers can design GenAI interfaces that encourage users to engage with the system in ways that promote collaboration rather than outsourcing. | • Policies should take biases on the side of the human interacting with GenAI into account, in the endeavor to create awareness of such biases and formulate specific guidelines to mitigate them.<br>• Policy makers could require tech developers to incorporate feedback systems that offer users accurate estimates of their own contribution to outputs co-produced with GenAI.<br>• Policy makers need to decide which types of usage behavior are (un)acceptable and motivate tech developers to design interfaces that make unacceptable behaviors difficult or impossible. | • Encouraging conversations about acceptable AI usage in social networks.<br>• Fostering a collective understanding of which usage behavior is acceptable and which biases may exist in this evaluation. |

*Transparency and disclosure of GenAI usage.* Disclosing GenAI usage may not be top of mind for users, who likely have become accustomed to using other tools (e.g., Google, spelling checkers, calculators) without disclosing the help those tools provided. If users construe GenAI as just another tool, they might perceive no need for disclosure. But encouraging users to spell out their exact usage behavior may help creators and evaluators form more realistic judgments of the extent of GenAI usage. We note the need for transparency and explainability with respect to GenAI and the data on which it is based but also issue a related call for transparency and explainability by users, with respect to their GenAI usage behavior. In line with this, some publishers already actively encourage such disclosure (though with varying levels of specificity), whereas others remain silent thus far.

As a final disclaimer, Table 2 illustrates the relevance of our work for a multitude of stakeholders but does not offer an exhaustive list of implications for all parties who encounter GenAI in their day-to-day lives. For example, intellectual property rights over creative work are particularly relevant for artists and musicians. Our experimental paradigms do not relate explicitly to such usage behaviors, but our findings might generalize to these instances, with the implication that artists and musicians perceive that they deserve substantial credit for artistic productions, whereas evaluators may underestimate their contributions.

### *11.2. Limitations and directions for future research*

With the hype surrounding ChatGPT it is not sufficient to focus only on the *tool*, but it is also important to put the *humans* interacting with the tool in the spotlight. Zooming in on people's subjective perceptions surrounding GenAI offers many exciting opportunities for future research, some of which we consider below. In particular, we identify different inferences of GenAI usage behavior as a driver of the self–other bias in inferred contribution. Yet, one may wonder which process(es) underlie these differential inferences of usage behavior. Turning to vast literature on self–other differences (e.g., Brown, 2012; Chambers & Windschitl, 2004), we anticipate that both cognitive and motivational factors are at play. Cognitively, creators know how they intended to use GenAI (i.e., for inspiration only), but those intentions are not known to evaluators. Motivationally, creators may downplay the extent to which they have used GenAI, in an attempt to protect their own self-image. To disentangle these processes and understand which one might be more influential, continued research could feature separate interventions, aimed at evoking each process. Cognitive-oriented interventions could explicitly highlight the positive intentions of creators or nudge evaluators to think about the full range of usage behaviors, to make the full spectrum of intentions salient (for a similar intervention, see Jung et al., 2020). A motivation-oriented intervention instead might frame GenAI usage as a behavior that reflects positively on the self (e.g., using it to complete mundane tasks efficiently). We consider it particularly relevant to explore boundary conditions that might shed light on the role and weight of different processes in driving the effect we identify.

In the realm of discussing self-enhancement motives, social desirability may seem a likely driver of the effect. Yet, there are two reasons for why we consider it unlikely to be the main driver of the bias that we observe. In Studies 2a and 2b, participants had to use ChatGPT for an unrelated task, prior to the focal task, which should have implied that ChatGPT usage was acceptable—and potentially even encouraged. More generally, the massive hype around ChatGPT usage, and the presence of millions of adopters, suggests that ChatGPT usage already is popular and "normal." Indeed, the study by Noy and Zhang (2023) documented that 80% of all participants openly revealed outsourcing their task to ChatGPT. Whereas these reasons make social desirability unlikely, we leave it to future research to explore such concerns around GenAI usage.

As a concluding remark, the expanding prevalence of GenAI in society is establishing a new reality, in which more and more daily tasks can be performed by or with technology. It is unrealistic to assume that people—students, authors, marketing professionals—will refrain from using GenAI tools, considering the many advantages they offer. Rather, the choice that people face is to outsource the task completely to a GenAI tool or to co-produce outcomes with it. Many stakeholders, including educational institutions and journals, may want to advocate explicitly for collaborative practices. Moving forward, this makes it essential to further illuminate people's perceptions and biases surrounding GenAI usage, as well as identify practices to eliminate biased perceptions.

## 12. Declaration of GenAI and AI-Assisted Technologies Used in the Writing Process

During the preparation of this work, the authors used GenAI (ChatGPT) to generate the stimuli for Study 2 (i.e., usage domains) and Study 4 (i.e., usage behaviors), as well as title suggestions. After using this tool/service, the authors reviewed and edited the content; they take full responsibility for the content of the publication. According to the living guidelines formulated in *Nature* (https://www.nature.com/articles/d41586-023-03266-1), the authors disclose their exact usage behavior of GenAI as follows:

1. GenAI was neither used for data interpretation nor for writing the manuscript.
2. GenAI was used for:
   a) Generating suggestions for titles based on the abstract of this work.
   b) Generating stimuli (i.e., usage domains) for Study 1.
   c) Generating stimuli (i.e., usage behaviors) for Study 3.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

d) Generating suggestions for code on R.
3. We have exclusively used ChatGPT (version GPT 3.5) for the tasks listed under 2.
4. We did not preregister the usage of ChatGPT, because it pertained only to the creation of stimuli, not to hypothesis development, testing, or analysis.
5. We did not extensively use a GenAI tool and thus did not have to replicate our findings using a different tool.

## CRediT authorship contribution statement

**Begum Celiktutan:** Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Anne-Kathrin Klesse:** Conceptualization, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing. **Mirjam A. Tuk:** Conceptualization, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing.

## Data availability

Data, R-code, materials, preregistrations, and IRB approvals are available at https://osf.io/qspwd/?view_only=3854679bd33c4699bf4dab4fcb22c4cb.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank Romain Cadario for his excellent comments on an earlier draft of this manuscript as well as Jordi Quoidbach for suggesting the follow-up analysis with an AI detector tool. We also thank the Erasmus Research Institute of Management (ERIM) for funding the data collection, Erik Kemperman from the Research Software Engineering and Consulting (RSEC) team at the Rotterdam School of Management, Erasmus University for his technical support with the data collection of study 2, and Danial Hayati for his help with collecting the data of Study 3.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijresmar.2024.05.006.

## References

Acar, O. A. (2023). A practical guide for marketers who want to use GenAI. *Harvard Business Review*. https://hbr.org/2023/12/a-practical-guide-for-marketers-who-want-to-use-genai.

Agarwal, S., De Freitas, J., Ragnhildstveit, A., & Morewedge, C. K. (2024). Acceptance of automated vehicles is lower for self than others. *Journal of Association for Consumer Research, 9*(3). in press.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*(6), 1621–1630. https://doi.org/10.1037/0022-3514.49.6.1621.

Barrick, E. M., Barasch, A., & Tamir, D. I. (2022). The unexpected social consequences of diverting attention to our phones. *Journal of Experimental Social Psychology, 101*, 1–14. https://doi.org/10.1016/j.jesp.2022.104344.

Barros, A., Prasad, A., & Śliwa, M. (2023). Generative artificial intelligence and academia: Implication for research, teaching and service. *Management Learning, 54*(5), 597–604.

Bergner, A. S., Hildebrand, C., & Häubl, G. (2023). Machine talk: How verbal embodiment in conversational AI shapes consumer–brand relationships. *Journal of Consumer Research, 50*(4), 742–764.

Bockting, C. L., van Dis, E. A., van Rooij, R., Zuidema, W., & Bollen, J. (2023). Living guidelines for generative AI—why scientists must oversee its use. *Nature, 622*(7984), 693–696.

Bonezzi, A., & Ostinelli, M. (2021). Can algorithms legitimize discrimination? *Journal of Experimental Psychology: Applied, 27*(2), 447–459.

Botti, S., & Iyengar, S. S. (2004). The psychological pleasure and pain of choosing: When people prefer choosing at the cost of subsequent outcome satisfaction. *Journal of Personality and Social Psychology, 87*(3), 312–326. https://doi.org/10.1037/0022-3514.87.3.312.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4*(4), 353–376. https://doi.org/10.1521/soco.1986.4.4.353.

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin, 38*(2), 209–219.

Bruk, A., Scholl, S. G., & Bless, H. (2018). Beautiful mess effect: Self–other differences in evaluation of showing vulnerability. *Journal of Personality and Social Psychology, 115*(2), 192–205.

Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour, 5*(12), 1636–1642.

Castelo, N., Boegershausen, J., Hildebrand, C., & Henkel, A. P. (2023). Understanding and improving consumer reactions to service bots. *Journal of Consumer Research, 50*(4), 848–863.

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

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130*(5), 813–838.

Chui, M., Yee, L., Hall, B., Singla, A., & Alexander, S. (2023). The state of AI in 2023: Generative AI's breakout year. *McKinsey*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730. https://doi.org/10.1037/0022-0663.88.4.715.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*, 24–42.

De La Garza, A. (2020). States' automated systems are trapping citizens in bureaucratic nightmares with their lives on the line. *Time.* https://time.com/5840609/algorithm-unemployment/.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., & Ahuja, M. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*.

Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology, 14*, 1199058.

Epley, N., & Dunning, D. (2000). Feeling" holier than thou": Are self-serving assessments produced by errors in self-or social prediction? *Journal of Personality and Social Psychology, 79*(6), 861–875.

Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in Education: Global Reactions to AI Innovations. *Scientific Reports, 13*(1), 1–14.

Gai, P. J., & Klesse, A.-K. (2019). Making recommendations more effective through framings: Impacts of user-versus item-based framings on recommendation click-throughs. *Journal of Marketing, 83*(6), 61–75.

Goldman, S. (2023). *Who owns DALL-E images? Legal AI experts weigh in*. VentureBeat. https://venturebeat.com/ai/who-owns-dall-e-images-legal-ai-experts-weigh-in/.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: The Guilford Press.

Howe, P. D. L., Fay, N., Saletta, M., & Hovy, E. (2023). ChatGPT's advice is perceived as better than that of professional advice columnists. *Frontiers in Psychology, 14*, 1281255.

Hutson, M. (2023). Rules to keep AI in check: Nations carve different paths for tech regulation. *Nature, 620*(7973), 260–263.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology, 76*(3), 349.

Jago, A. S., & Carroll, G. R. (2023). Who made this? Algorithms and authorship credit. *Personality and Social Psychology Bulletin*.

Jones, E., E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* New York: General Learning Press.

Jung, M. H., Moon, A., & Nelson, L. D. (2020). Overestimating the valuations and preferences of others. *Journal of Experimental Psychology: General, 149*(6), 1193.

Khan Academy. *World-class AI for education.* https://www.khanacademy.org/khan-labs.

Klesse, A.-K., Cornil, Y., Dahl, D. W., & Gros, N. (2019). The secret ingredient is me: Customization prompts self-image-consistent product perceptions. *Journal of Marketing Research, 56*(5), 879–893.

Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self-assessment: The road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin, 30*(3), 328–339.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480.

Longoni, C., Tully, C., & Shariff, A. (2023). Plagirizing AI-generated content is seen as less unethical and more permissible. *PsyArxiv*. https://doi.org/10.31234/osf.io/na3wb.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586.

Orrù, G., Piarulli, A., Conversano, C., & Gemignani, A. (2023). Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in Artificial Intelligence, 6*.

Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing, 40*(2023), 269–275.

Polman, E., Ziano, I., Wu, K., & Van Kerckhove, A. (2022). Consumers believe that products work better for others. *Journal of Consumer Research, 49*(1), 132–153.

Pronin, E. (2008). How we see ourselves and how we see others. *Science, 320*(5880), 1177–1180.

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

Roose, K. (2023). *Don't ban ChatGPT in schools*. The New York Times: Teach with it. https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.

Sun, D. (2023). How enterprises can regulate the development of generative AI. *Forbes.* https://www.forbes.com/sites/forbestechcouncil/2023/06/02/how-enterprises-can-regulate-the-development-of-generative-ai/?sh=1f9464f21626.

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313. https://doi.org/10.1126/science.adg7879.

Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*(7947), 224–226.

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*(1), 1–8.

Watson, D. (1982). The actor and the observer: How are their perceptions of causality divergent? *Psychological bulletin, 92*(3), 682–700. https://doi.org/10.1037/0033-2909.92.3.682.

Williams, E. F., Gilovich, T., & Dunning, D. (2012). Being all that you can be: The weighting of potential in assessments of self and others. *Personality and Social Psychology Bulletin, 38*(2), 143–154.

Williams, E. F., & Steffel, M. (2014). Double standards in the use of enhancing products by self and others. *Journal of Consumer Research, 41*(2), 506–525.

Yalcin, G., Themeli, E., Stamhuis, E., Philipsen, S., & Puntoni, S. (2023). Perceptions of justice by algorithms. *Artificial Intelligence and Law, 31*(2), 269–292.

Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making, 18*(e41), 1–16.

Zhang, Y., Tuk, M. A., & Klesse, A.-K. (2024). Giving AI a human touch: Highlighting human input increases the perceived helpfulness of advice from AI coaches. *Journal of Association for Consumer Research, 9*(3). in press.



Please cite this article as: B. Celiktutan, A.-K. Klesse and M.A. Tuk, Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations, International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2024.05.006