



Kent Academic Repository

Claessens, Scott, Veitch, Pierce and Everett, Jim A.C. (2025) *Negative perceptions of outsourcing to artificial intelligence*. Computers in Human Behavior . ISSN 0747-5632.

Downloaded from

<https://kar.kent.ac.uk/112415/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.chb.2025.108894>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



Negative perceptions of outsourcing to artificial intelligence

Scott Claessens , Pierce Veitch , Jim A.C. Everett ^{*}

School of Psychology, University of Kent, United Kingdom

ARTICLE INFO

Keywords:

Artificial intelligence
Person perception
Outsourcing
Effort
Trust

ABSTRACT

As artificial intelligence (AI) tools become increasingly integrated into daily life, people are beginning to outsource not only professional tasks but also socio-relational ones. Large language models like ChatGPT can generate wedding vows, speeches, and personal messages, raising questions about how individuals who use AI for such tasks are perceived by others. In this paper, we conduct six pre-registered studies with British participants ($N = 3935$) to understand how people view those who expend less effort by outsourcing tasks to AI in different ways. We highlight a tradeoff between efficiency and inferred moral character, authenticity, and value: outsourcing makes us think more negatively about not only the person and their motivations, but also the outsourced work itself. Importantly, this effect is not uniform. Reduced effort does not consistently lead to domain-general negative character perceptions across all tasks, but has particularly negative effects for outsourcing socio-relational tasks. Our results suggest that reduced effort is important not only because people value time and energy spent, but because expending less effort through outsourcing signals second-order perceptions that people are being less authentic and care less about the task. Our research highlights how relying on AI shapes our perceptions of the user, raising key philosophical questions about efficiency, authenticity, and social ties in a world filled with AI-mediated interactions.

The widespread release of generative AI language models has transformed daily life, offering the potential to perform a variety of tasks more efficiently and, in some cases, with greater effectiveness than by doing them oneself. But as AI becomes more widely available, people are not only using it to assist them with things like preparing dinner recipes, writing data analysis code, and planning daily schedules. Increasingly, AI might be used beyond routine or technical domains to assist in tasks that are more socio-relational in nature, like writing wedding vows, apology notes, and love letters. Anecdotal evidence suggests that not only is AI-outsourcing of this kind already happening, but that it potentially has serious effects on how we judge others. In a recent Reddit post, a disgruntled newlywed tells the story of her husband using ChatGPT to write his wedding vows, expressing her discomfort with outsourcing something to AI that, to her, is deeply meaningful and a reflection of their love for one another (miramar0, 2024). Other reports describe people's negative reactions when they learn that their romantic partner used ChatGPT to write them an apology note (Tait, 2024) or even to break up with them (Anderson, 2025). Against a backdrop of calls to increase trust and usage of AI, outsourcing tasks – especially socio-relational ones – to AI tools may be efficient, but this efficiency could trade off against inferred moral character, authenticity, and value.

In a world where the landscape of expending effort has changed through large language models (LLMs), we need to know how people who outsource tasks to AI are perceived – and why. We know that people are increasingly using LLMs for a wide variety of tasks (Department for Science, Innovation & Technology, 2024). Given the ubiquity of LLMs and their labelled intention as a tool for efficiency, it is possible that outsourcing to LLMs might not have consequences for perceived character. But we are sceptical. We know that people value products that they have put more effort into (Norton et al., 2012), that they dislike it when others “free ride” or reduce effort while benefiting from collective resources (e.g. Cubitt et al., 2011; Kerr, 1983), and that people's outputs are perceived as more valuable the more effort was ostensibly put into them (Kruger et al., 2004). And while there is much less attention paid in social psychology to how expending effort shapes perceptions of character, the evidence there suggests that exertion of effort is deemed morally admirable and is rewarded, even in situations where effort does not directly generate additional product, quality, or economic value, suggesting that effort itself is moralized (Celniker et al., 2023). While outsourcing and reduced effort are not synonymous, they are closely linked and the core social psychological processes are likely to be similar: when someone outsources a task instead of doing it themselves,

^{*} Corresponding author. Keynes College, Canterbury, CT2 7NP, United Kingdom.

E-mail address: j.a.c.everett@kent.ac.uk (J.A.C. Everett).

<https://doi.org/10.1016/j.chb.2025.108894>

Received 15 July 2025; Received in revised form 14 November 2025; Accepted 13 December 2025

Available online 17 December 2025

0747-5632/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

they are expending less time and energy to achieve their goal, and people value effort. Indeed, the scant recent work that does exist suggests that outsourcing tasks to LLMs could have negative consequences for person perception: describing someone as using AI for a relational task led to the perception they expended less effort and were less satisfied with their relationship (Liu et al., 2024) and people receive more negative social evaluations if they use AI to complete assignments at school and at work (Reif et al., 2025; Roth & Tissot, 2025). We do not know, however, how outsourcing – and therefore, presumably, expending less effort – might be different for a human or an AI, and how these processes will occur in the myriad ways in which people can and do use AI to help them complete tasks.

Second, we need to know how different ways of outsourcing to AI influence negative perceptions. There are many different ways that people can use LLMs, and someone who “fully” outsources a task to AI by simply giving it a prompt and copying the output word-for-word might be perceived very differently to someone who gives the AI a prompt, revises the work accordingly, and finishes it themselves – using AI as a collaborative tool, rather than as a replacement. Similarly, someone could deceive others about their use of AI or be perfectly honest about it, and someone could use AI because they are in a rush or because the task is important to them and they want to get it right. In all cases people are expending less effort, but the way they do this is different and it remains unclear how much this influences perceptions, and whether using AI in this way is sufficient to eliminate negative perceptions: if someone does not do the task fully themselves but uses an AI tool in the “best” way (by using it as a collaborative tool, being open about their use of AI, and having a good reason for using AI), would they still suffer negative social consequences from doing so?

Third, we need to know whether the *type* of task that people are outsourcing matters. One might expect outsourcing to be perceived negatively regardless of the type of task being outsourced – if effort is generally moralized, then the domain in which it is expended (or not) should have little impact. However, there are reasons to expect differences between socio-relational tasks that impact people’s relationships with others (such as writing wedding vows) and non-social tasks that are undertaken merely to achieve an instrumental goal (such as writing computer code). We know that different norms, standards, and expectations can be applied to social and non-social tasks and exchanges (e.g. A. P. Fiske, 1992; Heider, 1958; Malle, 2022) and that what people want and expect from different kinds of relationships (and, therefore, associated tasks) varies (Clark & Mills, 2012). Indeed, from a philosophical perspective, it often matters not only *whether* something is done, but *how* it is done (Aristotle, 2009; Hursthouse & Pettigrove, 2023; Stohr, 2006). An apology is not just about hearing someone say “I am sorry”, but seeing genuine regret; a love letter is not just about hearing someone say “I love you”, but seeing depth of emotion; and a bereavement letter is not just about hearing someone say “I am sorry for your loss”, but seeing an understanding for the powerful human experience of loss. There is, perhaps especially for social tasks, value not only in the outcome of doing something, but the *process* too (Goodman, 2010). Previous work on the role of effort, however, has seemed to suggest that effort should be uniformly important regardless of domain and has tended to focus on a limited number of tasks, looking at how people judge the quality of products like paintings, poems, or suits of armour based on how much effort is perceived (Kruger et al., 2004) or how people judge others who expend more or less effort in employment and physical effort like marathon running (Celniker et al., 2023). Yet the number of tasks that people can outsource – and therefore, presumably, expend less effort on – is vast, and these tasks can vary in different ways that seem likely to have very different consequences for person perception. Different kinds of tasks that people can outsource can be more or less socially relevant, more or less consequential, and more or less likely to impact others. Whether someone uses AI to spend less time planning a friend’s birthday party or writing an apology letter seems intuitively to be a different kind of thing than using AI to spend less time writing a daily schedule or

computer code to present a figure in an academic paper. To understand any potential negative effects of outsourcing – to AI or to humans – we must therefore look at a broad range of non-social and social tasks to understand not only potential generalizability but also *differences*.

Fourth, we need to know how outsourcing to AI, in different kinds of tasks and in different ways, may shape different *kinds* of social perceptions. Decades of research in social psychology have shown that people make inferences about the character of others along two key dimensions, broadly referred to as warmth and competence (e.g., Abele et al., 2021; S. T. Fiske et al., 2007). Broadly, competence refers to a person’s ability to obtain their desires, preferences or goals, while warmth is typically understood as referring to sociality and good intentions (though more recent work has explicitly distinguished warmth from morality and trustworthiness, Goodwin et al., 2014). While previous work has considered how lack of effort in more instrumental tasks shapes character judgments across these dimensions (Celniker et al., 2023), it remains unclear how outsourcing to AI in particular might lead to differential character judgments across these different dimensions (and across different tasks) nor how people might perceive laziness specifically (e.g., Ahmad et al., 2023).

Fifth, we need to understand *why* outsourcing to AI, and therefore expending less effort, might have these effects. Why is it that delegating to an AI might have negative consequences for person perception – especially if people are, in one sense, only using the AI tool in the precise way it has been designed and marketed? Previous work has focused on how expending less effort directly leads to negative perceptions of others because effort itself is moralized and seen as a cue of cooperative intent (Celniker et al., 2023). But – especially if there are differences in how people are judged from outsourcing and from which tasks – this still leaves open the question of *why* effort is seen as important and what exactly it is signalling to others, beyond one’s general cooperative intent. If expending less effort through outsourcing to AI has different consequences for different character judgments across different tasks, this could suggest that it is not only that effort itself is important, but that a lack of effort spent on the task signals more broadly about the person, in line with more virtue-ethics based discussions in philosophy.

As well as outsourcing to AI indicating less effort expended and this directly driving character judgments, it is possible that when someone outsources to AI they are seen as producing work that is less authentic, and this drives negative character perceptions. Authenticity – understood as “true expression of an individual’s [...] values” (Dutton, 2003, p. 259; see also Newman, 2019) – has often been discussed in the context of psychological wellbeing (Sutton, 2020) and leadership (Banks et al., 2016; Gardner et al., 2011) and has been argued to play an important role in various aspects of healthy psychological functioning and interpersonal relations (Kernis & Goldman, 2006). But we also know that perceived authenticity shapes how we value things. Authenticity has been highlighted as a reason why people value original artwork over identical duplicates (or forgeries), because people think about art objects as emerging from a unique creative act and typically requiring physical contact (Newman & Bloom, 2012). Discussion of authenticity has only become more pressing in the age of AI, with the 2023 Merriam Webster word of the year being “authenticity” in part due to the increased use of LLMs like ChatGPT (Italie, 2023), and scholars have begun to consider how perceived authenticity shapes how people view AI service apps (Vo et al., 2025). But how would this influence character judgments for users of AI? In their humanist work on the transparent self, Jourard (1971) argued that “authentic being means being oneself, honestly, in one’s relations with his fellows” (p. 133). If someone uses an AI tool to complete tasks for them, especially socio-relational tasks, they may well be perceived as being inauthentic and this may shape how people perceive those who outsource to AI.

Another way that outsourcing to AI could lead to negative perceptions, distinct from effort itself, is perceptions of how much someone cares about the task at hand. Celniker et al. (2023) focus on how effort serves as a signal of one’s cooperative intent and therefore moral

character, but it is also possible that, more specifically, the precise task that someone expends effort on is taken as a signal of how much they care about that task and what it represents. Self-determination theory (e.g., Ryan & Deci, 2000) highlights how people are more likely to pursue and attain a goal when it meets their core psychological motives, and it is possible that people intuitively infer this when evaluating others and their values. For example, a student who puts more effort into their schoolwork may be perceived more positively for this if effort is moralized, but they may also just be perceived as *caring more* about their academic development. Indeed, recent evidence suggests that people do infer perceptions of meaningfulness from effort, for themselves and others (Campbell et al., 2025). In a series of studies, Campbell et al. (2025) explore how expending effort can increase felt meaning so that, for example, when participants completed a writing task manually instead of expending less effort by using ChatGPT, they perceived their own work as more meaningful. Moreover, these inferences also appear to apply to others, with one study showing that participants expected that someone who spent more time on an academic assignment found it more meaningful. While focused on meaning, not authenticity or valuation per se, these results suggest expending effort can have second-order effects on how people are perceived. It remains unclear, however, how such second-order perceptions would drive different kinds of character judgments from outsourcing to AI specifically (and outsourcing to AI in different ways), nor how this might vary based on the different kinds of tasks that one could outsource.

1. Present research

In this paper, we explore how people perceive others who outsource different kinds of tasks, in different ways, for different reasons, to AI. Across six pre-registered experiments with British participants, we seek not only to understand how reduced effort through AI-outsourcing might shape perceptions of others but also advance previous theoretical work on the moralization of effort (Celniker et al., 2023). Moving beyond a focus on how effort – typically in non-social instrumental tasks – shapes character judgments uniformly across different domains, we seek to provide deeper theoretical nuance into why effort is important: that effort might be particularly important for socio-relational tasks, not only because effort itself is intrinsically moralized, but because expending effort gives rise to second-order inferences about authenticity and care that independently influence character judgments.

In Study 1, we explore the perceptions of people who complete a task themselves, outsource to another person, or outsource to AI, using a large set of tasks with different levels of social relevance (e.g., from writing daily schedules, computer code, and dinner recipes to wedding vows, apology letters, and bereavement cards). In Study 2, we then look at the effects of task type, AI use, and honesty, again looking across our large range of tasks and manipulating whether people use AI as a collaborative tool or “fully” outsource to AI and whether they are honest or deceptive about their use of AI. After turning to look at perceptions of both outsourcers and the outsourced work in Study 3, in Studies 4–6 we probe why outsourcing may have negative effects on how we evaluate others. In Study 4, we tease apart potential mechanisms of effort spent and authenticity by looking at how people evaluate others who either spend a lot or little time crafting the AI prompts, and who either outsource to a generic or personalized AI. In Study 5, we test the potential mechanism of perceived importance in the task by manipulating people’s reasons for using AI – either because they wanted to save time or because they cared about the task and thought that AI would improve their work. Finally, in Study 6, we bring these different potential mechanisms together to explore the different pathways that influence the relationship between outsourcing and negative perceptions, exploring perceptions of effort spent, how authentic the person was, and how much the person was perceived to value the task itself. Table 1 summarises the different theoretical constructs focused on in all six studies.

Table 1

Key constructs and operationalisations across studies.

Construct	Meaning	Operationalisation
Outsourcing	The extent to which completion of a task is delegated to an external agent (human or AI), rather than performed by the individual themselves.	An experimental manipulation in which - in comparison to the control condition where a person ‘works on it by themselves from start to finish’ (Control) - the person either ‘gets someone else to do it for them’ (Human outsourcing: Study 1) or ‘gets the AI tool ChatGPT to do it for them’ (AI Outsourcing: Study 1).
Fully outsourcing to AI	The extent to which someone fully delegates the completion of a task to an AI.	An experimental manipulation in which we describe how the person completes the task: ‘In order to complete this task, [person] uses the AI tool ChatGPT. They copy ChatGPT’s output word-for-word, rather than doing it themselves’ (Studies 2–3, 5–6).
Outsourcing to AI as a tool	The extent to which someone uses AI as a tool for assistance, but still completes the task themselves.	An experimental manipulation in which we describe how to the person completes the task: ‘In order to complete this task, [person] uses the AI tool ChatGPT. They ask ChatGPT to provide ideas, inspiration, and feedback, but they edit and rewrite the suggestions and finish the task themselves’ (Studies 2–3, 5–6).
Effort	The (perceived) degree of time and energy to complete a task.	Directly measured as a dependent measure tracking people’s perception (Study 6): ‘How much effort did [person] put into the love letter?’ Directly manipulated through description of time and energy extended (Study 4), where someone ‘carefully gives the AI several detailed prompts and, after multiple rounds of changes, uses its resulting output’, or puts in less effort when they ‘quickly give the AI a rushed prompt and use its first output.’ Indirectly manipulated through the manipulation of someone completing a task themselves (highest effort), using AI as a tool but finishing it themselves (medium effort), or fully outsourcing to AI (lowest effort) (Studies 2–3, 5–6).
Authenticity	The (perceived) extent to which something is personal or a true expression of their values.	As a dependent measure tracking people’s perceptions (Study 6): ‘How authentic do you think [person’s] love letter was?’ As a manipulation (Study 3) in which the person used a personalized LLM trained on their own writing, compared to a generic LLM.
Care	The (perceived) extent to which someone cares about and values the task.	As a dependent measure tracking people’s perceptions (Studies 5, 6): ‘How much do you think [person] cares about this task?’ As a manipulation (Study 5) in which the person is described as using AI ‘because this task is really important to them and they want to make sure they get it right’, or instead using it because ‘they are really short on time and want to complete the task quickly’.

2. Study 1

2.1. Methods

2.1.1. Ethical approval

Ethical approval was granted for all studies in this paper by the University of Kent Psychology Research Ethics Panel. Participants in all studies provided informed consent and were debriefed after the study.

2.1.2. Participants

We conducted a power simulation to determine our target sample size (see Supplementary Methods). The simulation suggested that a sample size of 150 participants per condition (overall $n = 450$ for three conditions) would be required to detect a small difference between conditions (Cohen's $d \approx 0.20$) with above 80% power.

We recruited a convenience sample of 500 participants from the United Kingdom through the online platform Prolific (<https://www.prolific.com/>). After excluding participants who failed our pre-treatment attention check, we were left with a final sample of 466 participants (292 female; 169 male; 4 non-binary/third gender; 1 undisclosed gender; mean age = 42.32 years). 73% of these participants reported having used ChatGPT before.

2.1.3. Design

We randomly allocated participants into one of three conditions in a between-subjects design: (i) the control condition, (ii) the AI outsourcing condition, or (iii) the human outsourcing condition. These conditions determined how scenarios were presented to participants.

2.1.4. Procedure

We presented participants with six scenarios. Each scenario described a person completing a task, such as writing computer code or writing a love letter. The six tasks were randomly drawn from a larger set of 20 tasks (see Supplementary Table 1 for the full list of tasks). For each scenario, we told participants:

- *Control condition*: "In order to complete this task, [the person] works on it by themselves from start to finish."
- *AI outsourcing condition*: "In order to complete this task, [the person] gets the AI tool ChatGPT to do it for them."
- *Human outsourcing condition*: "In order to complete this task, [the person] gets someone else to do it for them."

We then asked participants how well each of the following words described the person in the scenario: competent, warm, moral, lazy, and trustworthy. Participants answered these questions on 7-point Likert scales, ranging from "does not describe [the person] well" to "describes [the person] extremely well".

After the six scenarios, we asked participants several questions about the AI tool ChatGPT, including their familiarity with ChatGPT, whether they had used ChatGPT before, how frequently they used ChatGPT, and how trustworthy they thought ChatGPT was.

2.1.5. Pre-registration

We pre-registered the study on the Open Science Framework (<https://osf.io/xhmzk>).

2.1.6. Statistical analysis

We fitted Bayesian multivariate multilevel cumulative-link ordinal models to the data using the *brms* R package. We used this modelling approach as the outcome variables were ordinal Likert scales and treating these as continuous can inflate false positive rates (Liddell & Kruschke, 2018). Since participants responded to multiple character evaluations at once, we modelled the character evaluations – competence, warmth, morality, laziness, and trustworthiness – as multivariate outcomes and included fixed effects for conditions, correlated varying

intercepts for participants, and correlated varying intercepts and slopes for tasks. We used regularizing priors for all parameters to impose conservatism on parameter estimates (see Supplementary Methods for full model specification, Supplementary Fig. 1 for posterior predictive checks, and Supplementary Table 2 for prior sensitivity check). All models converged normally ($\hat{R} \leq 1.01$; all effective sample sizes >400 ; no divergent transitions).

2.1.7. Transparency and openness

For all studies in this paper, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the studies. All studies were pre-registered. Analyses for all studies were conducted in R v4.4.2 (R Core Team, 2022). Visualizations were produced using the *ggplot2* and *patchwork* packages (Pedersen, 2025; Wickham, 2016). The manuscript was reproducibly generated using the *targets* package (Landau, 2021) and *quarto* (Allaire et al., 2024). All survey materials, data, and code to reproduce the analyses and figures in this paper can be found here: <https://osf.io/xhmzk>.

2.2. Results

We found that, overall, people who outsourced tasks to AI or other humans were perceived more negatively than people who completed the tasks themselves (Fig. 1). In particular, people who outsourced were perceived as lazier and less competent, with smaller yet detectable differences for perceptions of warmth, morality, and trustworthiness (Table 2). Across all measures, outsourcing to other humans was perceived more negatively than outsourcing to AI.

We found that the effects of outsourcing varied across the different tasks, especially for perceptions of warmth and morality (Fig. 2). For example, people were perceived as less warm if they outsourced writing a love letter, but not if they outsourced writing computer code. Similarly, people were perceived as less moral if they outsourced writing an apology letter to a friend, but not if they outsourced writing a dinner recipe. By contrast, the effects of outsourcing on competence, laziness, and trustworthiness were more consistent across tasks.

To determine the factors that predict variation across tasks, we incorporated ratings of tasks from a pilot study (see Supplementary Materials). Participants were asked to rate the 20 tasks on several features: whether the task is social, requires social skills, impacts others, has important consequences, and requires effort. All of these features predicted stronger causal effects of outsourcing compared to control (Supplementary Figs. 2 and 3). In other words, outsourcing to AI or other humans was perceived more negatively for tasks that were rated as being socially relevant and impactful.

2.3. Discussion

In Study 1, we conducted our first exploration of how outsourcing to either a human or an AI shaped character perceptions across a variety of tasks. Our results show that people who outsourced the task instead of completing it themselves were seen as less moral, warm, competent, and trustworthy, and more lazy. We found that outsourcing a task led to negative character judgments regardless of whether the task was outsourced to a human or a machine, though predictably we found that outsourcing to a human was seen as worse given that LLMs are at least purposefully designed to be used. And finally, we found that the kind of task matters: these negative perceptions were not uniform, but were particularly found for tasks that were more social and impactful.

In Study 2, we turned to focus specifically on the different ways that people might expend less effort by outsourcing to an AI, focusing on the role of type of use (using AI as a tool but still finishing the task oneself vs. full delegating) and honesty about the use of AI. Would these negative perceptions be seen even in the "best case" scenario where someone uses AI as a tool and is honest about it?

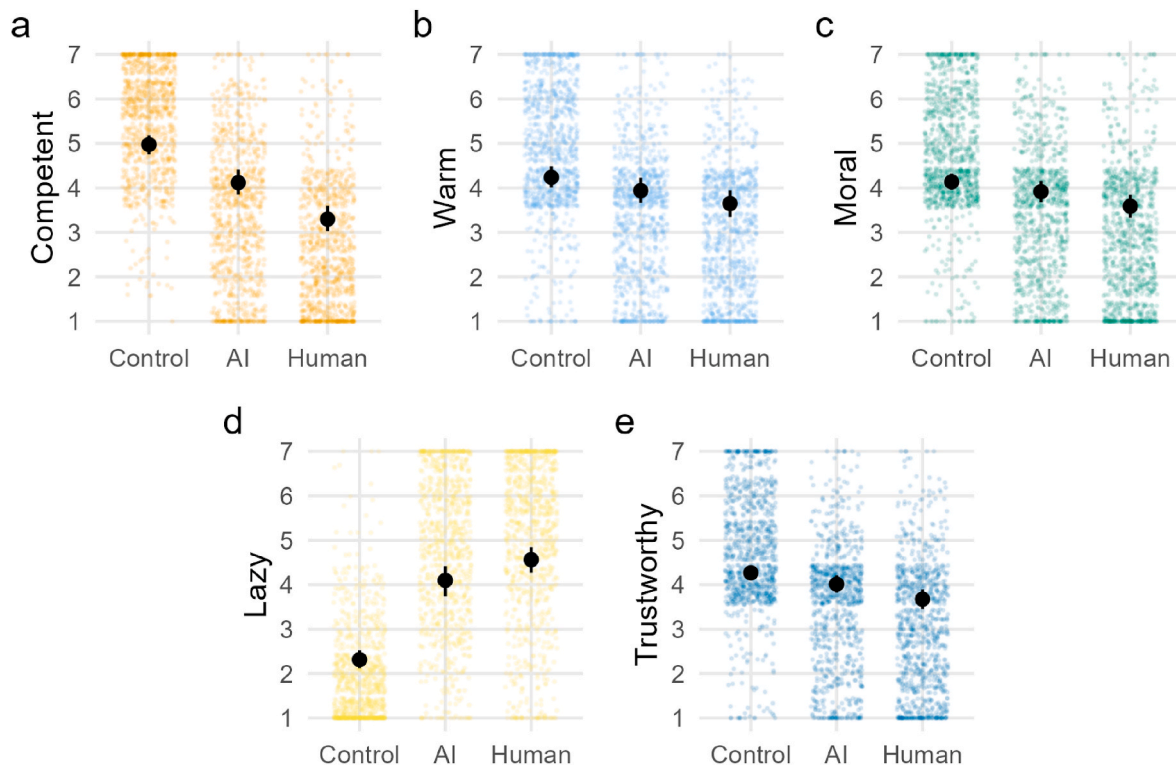


Fig. 1. Overall character evaluations in Study 1. *Note.* Participants in the control condition, the AI outsourcing condition, and the human outsourcing condition evaluated people in the scenarios on (a) competence, (b) warmth, (c) morality, (d) laziness, and (e) trustworthiness. Colored points represent participant responses to the questions, jittered for easier viewing. Black points are estimated marginal means from the fitted model, pooling over participants and tasks. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.

Table 2

Overall pairwise contrasts in Study 1.

	Response				
	Competent	Warm	Moral	Lazy	Trustworthy
AI - Control	−0.86 [−1.14 −0.56]	−0.30 [−0.57 −0.03]	−0.22 [−0.46 0.02]	1.78 [1.42 2.13]	−0.25 [−0.44 −0.06]
Human - Control	−1.68 [−1.97 −1.36]	−0.59 [−0.88 −0.29]	−0.55 [−0.81 −0.28]	2.25 [1.90 2.59]	−0.59 [−0.82 −0.37]
Human - AI	−0.82 [−1.16 −0.47]	−0.29 [−0.61 0.04]	−0.33 [−0.63 −0.03]	0.46 [0.07 0.89]	−0.34 [−0.59 −0.09]

Note. Numbers reflect differences in marginal means on a 7-point Likert scale, pooling over participants and tasks. Main numbers are posterior medians, numbers in the square brackets are 95% credible intervals.

3. Study 2

3.1. Methods

3.1.1. Participants

We used the same power estimate from Study 1 to determine our target sample size of $n = 750$ (150 participants in each of five conditions). We recruited a convenience sample of 800 participants from the United Kingdom through Prolific. After excluding participants who failed our pre-treatment attention check, we were left with a final sample of 762 participants (438 female; 316 male; 4 non-binary/third gender; 4 undisclosed gender; mean age = 42.16 years). 78% of these participants reported having used ChatGPT before.

3.1.2. Design

We used a “control plus 2x2” between-subjects design. Participants were randomly allocated to either the control condition, in which people in the scenarios complete the tasks themselves, or one of four experimental conditions, in which people in the scenarios use AI to complete the tasks. In the experimental conditions, we manipulated whether people in the scenarios used AI as a collaborative tool or “fully”

outsourced to AI, and whether people were honest or deceptive about their use of AI. This resulted in five conditions overall: (i) the control condition, (ii) the tool-honest condition, (iii) the tool-deception condition, (iv) the full-honest condition, and (v) the full-deception condition.

3.1.3. Procedure

The procedure was mostly identical to Study 1, but we focused on a subset of 16 tasks (see [Supplementary Table 1](#)) and told participants the following for each scenario:

- *Control condition:* “In order to complete this task, [the person] works on it by themselves from start to finish.”
- *Tool outsourcing conditions:* “In order to complete this task, [the person] uses the AI tool ChatGPT. They ask ChatGPT to provide ideas, inspiration, and feedback, but they edit and rewrite the suggestions and finish the task themselves.”
- *Full outsourcing conditions:* “In order to complete this task, [the person] uses the AI tool ChatGPT. They copy ChatGPT’s output word-for-word, rather than doing it themselves.”

We then told participants in the experimental conditions:

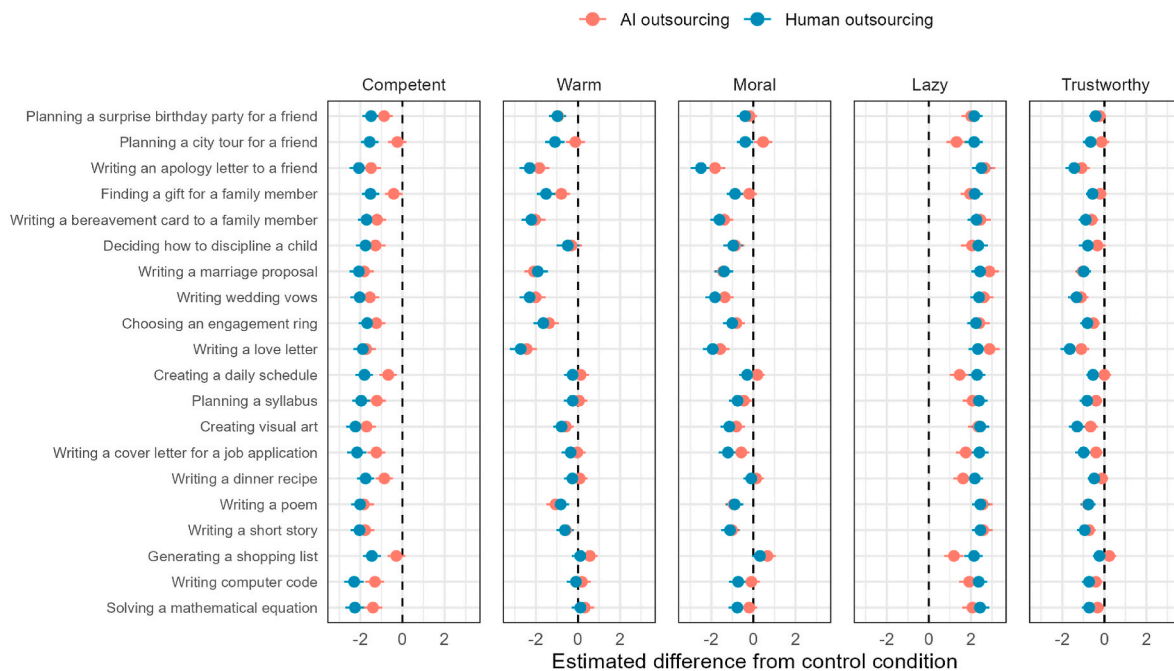


Fig. 2. Variation in the effects of outsourcing across tasks in Study 1. *Note.* Tasks are ordered from most social (top) to least social (bottom) according to ratings from the pilot study. Point ranges are differences in marginal means on a 7-point Likert scale for the AI outsourcing condition (red) and the human outsourcing condition (blue) compared to the control condition. Points and ranges represent posterior medians and 95% credible intervals, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

- **Honest conditions:** “After completing the task, [the person] is asked how they came up with their ideas. [The person] acknowledges that they used ChatGPT as a tool/got ChatGPT to do the task for them.”
- **Deception conditions:** “After completing the task, [the person] is asked how they came up with their ideas. [The person] does not acknowledge that they used ChatGPT as a tool/got ChatGPT to do the task for them.”

3.1.4. Pre-registration

We pre-registered the study on the Open Science Framework (<https://osf.io/xhmzk>).

3.1.5. Statistical analysis

We fitted the same Bayesian multivariate multilevel cumulative-link ordinal models as in Study 1 (see [Supplementary Fig. 4](#) for posterior predictive checks). All models converged normally ($\hat{R} \leq 1.01$; all effective sample sizes >400; no divergent transitions).

3.2. Results

We first looked at the overall results averaging over tasks. Across all five character evaluations, we found that fully outsourcing to AI (i.e., copying the AI output verbatim) was perceived more negatively than using AI as a collaborative tool ([Fig. 3](#); [Table 3](#)). By contrast, we found that deception about AI usage had specific negative effects on perceptions of morality and trustworthiness: people who did not acknowledge their use of AI were perceived as less moral and less trustworthy. Due to uncertainty in the parameter estimates, we were unable to draw conclusions about interaction effects between full outsourcing and deception.

The effects of outsourcing to AI varied across the different tasks, especially for perceptions of warmth and morality ([Fig. 4](#)). For example, people who used AI for social tasks, such as writing an apology letter or a bereavement card, were perceived as less warm, less moral, and lazier compared to people who completed the task themselves. This was true even if the person used AI as a tool and was honest about their usage of

AI. By contrast, we observed weaker effects of outsourcing for non-social tasks like writing computer code or planning a syllabus, highlighting again that reduced effort through outsourcing does not have uniform effects across tasks but it matters *what* people are outsourcing.

To determine the factors that predict variation across tasks, we again incorporated ratings of tasks from the pilot study. We found stronger causal effects of outsourcing to AI compared to the control for tasks that were rated as more social, requiring more social skills, impacting others more, having more important consequences, and requiring effort ([Supplementary Figs. 5 and 6](#)).

3.3. Discussion

In Study 2, we looked at how people who outsourced to AI in different ways were perceived across a broad range of social and non-social tasks. In line with our predictions, we found that “fully” outsourcing to AI was perceived more negatively than using AI as a collaborative tool, but these effects were not uniform across all character judgments and were seen particularly for socio-relational tasks. We also found, predictably, that people were seen as less moral and less trustworthy if they did not acknowledge their use of AI. Importantly, though, we show that even using AI in the “best” way – only as a tool and being honest about one’s usage – still led to negative social perceptions for the more socio-relational tasks like writing a love letter, an apology, or wedding vows.

In Study 3, we investigate whether these negative perceptions extend to the work itself and remain after seeing the output. It could be, for example, that someone is perceived badly for using ChatGPT to write their bereavement card, but the writing itself is seen as equally well-written and authentic, if not more so, than if the person had written the card themselves. Indeed, evidence suggests that text generated by ChatGPT is rated as higher quality than human-written text ([Noy & Zhang, 2023](#)). Moreover, it is possible that seeing appropriate output could mitigate negative perceptions by highlighting how the AI can in fact perform the task well. We explored these possibilities in Study 3.

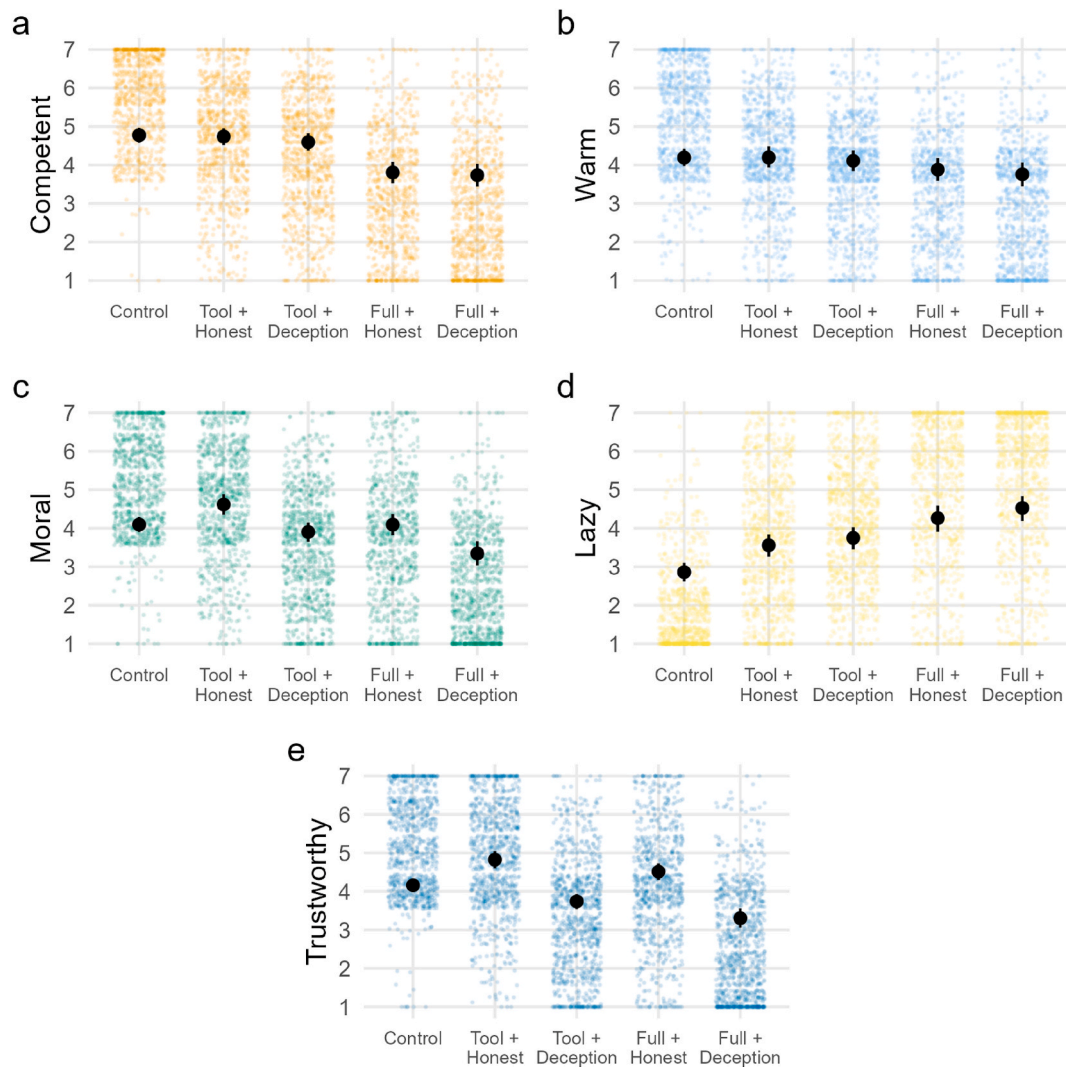


Fig. 3. Overall character evaluations in Study 2. *Note.* Participants in the control condition and four AI outsourcing conditions evaluated people in the scenarios on (a) competence, (b) warmth, (c) morality, (d) laziness, and (e) trustworthiness. Colored points represent participant responses to the questions, jittered for easier viewing. Black points are estimated marginal means from the fitted model, pooling over participants and tasks. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.

4. Study 3

4.1. Methods

4.1.1. Participants

We conducted a power simulation to determine our target sample size (see Supplementary Methods). The simulation suggested that a sample size of 125 participants per condition (overall $n = 750$ for six conditions) would be required to detect a small-to-medium difference between conditions (Cohen's $d \approx 0.40$) with above 80% power.

We recruited a convenience sample of 800 participants from the United Kingdom through Prolific. After excluding participants who failed our pre-treatment attention check, we were left with a final sample of 766 participants (425 female; 337 male; 3 non-binary/third gender; 1 undisclosed gender; mean age = 41.93 years). 72% of these participants reported having used ChatGPT before.

4.1.2. Design

We randomly allocated participants to one of six conditions in a 3x2 between-subjects design. We manipulated the type of outsourcing: (i) no outsourcing control, (ii) using AI as a tool, and (iii) fully outsourcing to AI. Here, in contrast to Studies 1 and 2, we also explicitly manipulated

whether the task prompt was social or non-social.

4.1.3. Procedure

We told participants that they would read and evaluate a short piece of writing from “another participant”. In reality, we generated the writing using ChatGPT version 4o. We asked ChatGPT to generate a 300 word response to the prompt and to write convincingly like a real human. We then edited the text to appear more human-like by, for example, removing classic AI markers like dashes and concluding sentences and ensuring that the information was not too generic, such that the writing could reasonably be attributed to both a human and AI.

The prompt for the piece of writing varied between conditions:

- *Social conditions:* “Please write a description of a close family member or friend, explaining why they are special to you.”
- *Non-social conditions:* “Please write a short description of a book, TV show, or film of your choice.”

We explained that the “other participant” was asked several questions about how they produced their answer, including whether or not they used an AI tool like ChatGPT. We explained that the participant was encouraged to be honest and told that they would be paid regardless.

Table 3

Overall pairwise contrasts in Study 2.

	Response				
	Competent	Warm	Moral	Lazy	Trustworthy
Comparison to control					
Tool Honest - Control	−0.03 [−0.28 0.20]	0.00 [−0.23 0.24]	0.52 [0.27 0.76]	0.70 [0.36 1.02]	0.66 [0.44 0.90]
Tool Deception - Control	−0.18 [−0.41 0.06]	−0.09 [−0.32 0.15]	−0.19 [−0.44 0.05]	0.89 [0.55 1.21]	−0.42 [−0.63 −0.21]
Full Honest - Control	−0.97 [−1.24 −0.70]	−0.31 [−0.56 −0.07]	0.00 [−0.26 0.26]	1.40 [1.02 1.76]	0.35 [0.14 0.59]
Full Deception - Control	−1.04 [−1.33 −0.74]	−0.44 [−0.70 −0.16]	−0.76 [−1.04 −0.45]	1.67 [1.30 2.04]	−0.86 [−1.12 −0.59]
Effect of full outsourcing					
Full Honest - Tool Honest	−0.93 [−1.24 −0.62]	−0.32 [−0.62 −0.01]	−0.52 [−0.86 −0.19]	0.71 [0.28 1.13]	−0.31 [−0.59 −0.03]
Full Deception - Tool Deception	−0.86 [−1.21 −0.52]	−0.35 [−0.67 −0.02]	−0.56 [−0.91 −0.21]	0.78 [0.37 1.20]	−0.44 [−0.73 −0.14]
Effect of deception					
Tool Deception - Tool Honest	−0.14 [−0.42 0.13]	−0.09 [−0.39 0.21]	−0.71 [−1.02 −0.40]	0.19 [−0.20 0.58]	−1.09 [−1.36 −0.80]
Full Deception - Full Honest	−0.07 [−0.43 0.29]	−0.12 [−0.45 0.21]	−0.75 [−1.10 −0.38]	0.26 [−0.16 0.71]	−1.21 [−1.51 −0.92]
Interaction effect					
Interaction effect	0.08 [−0.38 0.52]	−0.04 [−0.47 0.41]	−0.04 [−0.50 0.45]	0.07 [−0.50 0.67]	−0.12 [−0.54 0.28]

Note. Numbers reflect differences in marginal means on a 7-point Likert scale, pooling over participants and tasks. The bottom row represents the interaction between full outsourcing and deception (i.e., the difference between the differences in the rows above). Main numbers are posterior medians, numbers in the square brackets are 95% credible intervals.

The response from the “other participant” varied between conditions:

- **Control conditions:** “The participant reported that they did not use any AI tool like ChatGPT. Instead, they worked on the response themselves from start to finish.”
- **Tool outsourcing conditions:** “The participant reported using ChatGPT to provide ideas, inspiration, and feedback. The participant told us that they edited and rewrote ChatGPT’s suggestions and finished writing the response themselves.”
- **Full outsourcing conditions:** “The participant reported using ChatGPT to complete the task. The participant told us that they copied ChatGPT’s output word-for-word, rather than producing the response themselves.”

Next, we presented participants with a randomly-chosen pre-generated essay answer to the prompt (see [Supplementary Tables 3 and 4](#) for full essay answers). In the social conditions, the answer either referred to the participant’s father, their sister, or their best friend. In the non-social conditions, the answer either referred to the book *The Hobbit*, the TV show *Buffy the Vampire Slayer*, or the film *Titanic*. Reading times and responses to a follow-up comprehension question suggested that participants read the essay answers in sufficient detail (see [Supplementary Table 5](#)).

Finally, we asked participants about their perceptions of the essay answer and the “other participant”. We asked how well-written, meaningful, and authentic they thought the answer was (7-point Likert scales), what letter grade they would give the answer (A-E), and how much they would hypothetically reward the other participant for their work (from £0.00 to £1.00). We also asked how well each of the following words described the other participant: competent, warm, moral, lazy, and trustworthy (7-point Likert scales).

At the end of the study, we gave participants a manipulation check and asked them whether they believed the manipulation. Almost all participants correctly reported the condition that they were in and most participants stated that they believed the essay response was written in the way we described, suggesting that the manipulation was successful (see [Supplementary Table 6](#)). We also asked participants several questions about ChatGPT.

4.1.4. Pre-registration

We pre-registered the study on the Open Science Framework.¹

4.1.5. Statistical analysis

We fitted two Bayesian multilevel models to the data. The first model was a multivariate cumulative-link ordinal model including all Likert scales as separate response variables. The second model was a zero-one-inflated-beta model applied specifically to the reward question, which was a slider scale varying between 0 and 1. For both models, we included fixed effects for the interaction between outsourcing type and task type and varying intercepts and slopes for essay answers. We used regularizing priors for all parameters to impose conservatism on parameter estimates (see [Supplementary Methods](#) for full model specification and [Supplementary Fig. 7](#) for posterior predictive checks). All models converged normally ($\hat{R} \leq 1.01$; all effective sample sizes >400 ; no divergent transitions).

4.2. Results

We first looked at character evaluations. We found that even when provided with concrete output, people were still perceived more negatively across all character evaluations if they outsourced the writing task to AI, either by using ChatGPT as a collaborative tool or by copying the text from ChatGPT verbatim ([Supplementary Fig. 8](#); [Supplementary Table 7](#)). In contrast to Study 2, however, our analyses did not provide evidence of differences in character evaluations between the tool outsourcing and full outsourcing conditions. Due to uncertainty in the parameter estimates, we were unable to draw conclusions about differences between social and non-social tasks and interaction effects.

Turning to evaluations of the work itself, we found that the AI-outsourced work (either outsourced by using AI as a collaborative tool or fully outsourced) was judged as being equally well written to the work in the control condition ([Fig. 5](#); [Table 4](#)). This is in line with the writing indeed being identical in all conditions. Interestingly, however, we found that essay responses that were ostensibly generated using AI were perceived as less meaningful and less authentic compared to essay responses ostensibly written by a human. Participants also marked AI-generated essays with a lower grade and rewarded AI-generated essays with a lower hypothetical monetary bonus. In contrast to Study 2, we found no evidence of differences in perceptions of the work between

¹ Due to a technical error with archiving this pre-registration on the Open Science Framework, the timestamp for the registration was lost. However, on our OSF project (<https://osf.io/xhmzk>), it is possible to view our pre-registration document file and its timestamped upload date.

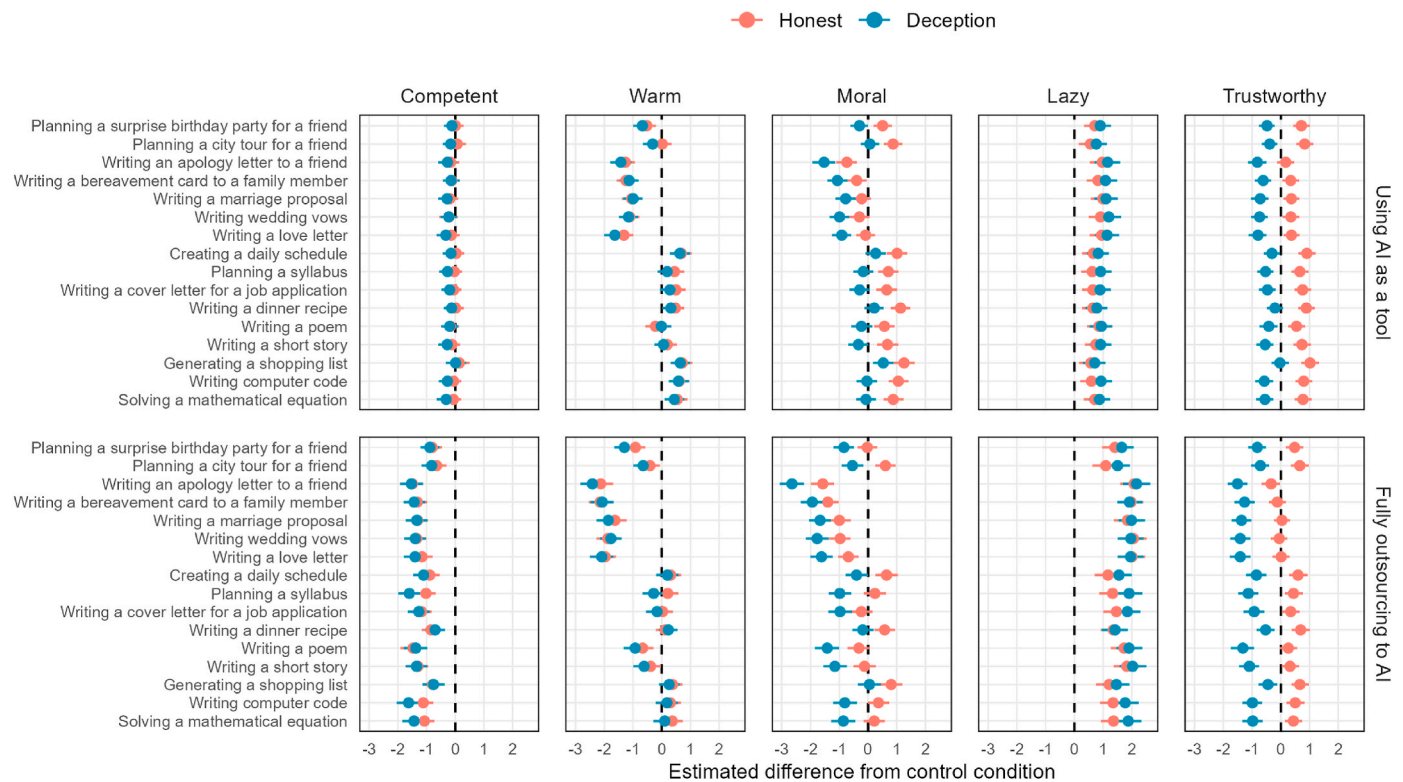


Fig. 4. Variation in the effects of outsourcing across tasks in Study 2. *Note.* Tasks are ordered from most social (top) to least social (bottom) according to ratings from a pilot study. Point ranges are differences in marginal means on a 7-point Likert scale for the honest conditions (red) and deception conditions (blue) compared to the control condition. Upper panels refer to the tool outsourcing conditions, and lower panels refer to the full outsourcing conditions. Points and ranges represent posterior medians and 95% credible intervals, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the tool outsourcing and full outsourcing conditions, except for the reward question, where fully outsourced essays (i.e., essays copied verbatim from ChatGPT) were rewarded £0.23 less than essays generated using AI as a collaborative tool. Due to uncertainty in the parameter estimates, we were unable to draw conclusions about differences between social and non-social tasks and interaction effects.

4.3. Discussion

In Study 3, we turned to look at how people perceived both the outsourcer and the outsourced work when given specific output in a social or non-social task that was described as being produced independently by a person, produced by a person in collaboration with AI as a tool, or outsourced in full to AI. We find that our results generalize from character judgments to perceptions of the work itself: text purportedly generated using AI was perceived to be less meaningful, less authentic, and less reward-worthy compared to the same text described as human-generated. We found no evidence of differences between social and non-social tasks, or between the tool and full outsourcing conditions, aside from the lower rewards given to participants in the latter condition. These null results are difficult to interpret, given the posterior uncertainty in the estimates.

In Study 4, we turn to explore potential mechanisms driving our effects. We assume that effort may play a role, since perceived effort is often used as a signal of one's moral character (Cubitt et al., 2011) and cooperative intent (Celniker et al., 2023). Study 3 also suggested a role of authenticity: in line with work on the psychological importance of authenticity (Newman, 2019), people who outsource to AI may be perceived as producing work that is less authentically their own, leading to negative evaluations. To what extent would negative perceptions of outsourcing persist when someone is described as putting a lot of effort

into the prompts that they fed to the LLM, and when someone is described as using a personalized – and therefore more “authentic” – LLM rather than a generic LLM like ChatGPT? To explore this, in Study 4 we experimentally manipulate (1) how much effort someone puts into the task and (2) whether they outsource the task to a standard LLM like ChatGPT or a personalized LLM trained specifically on their own prior writings (and so therefore producing work that is more authentically “theirs”). We expected negative perceptions of outsourcing to be mitigated when the person uses a personalized LLM and expends significant effort on formulating prompts for the AI.

5. Study 4

5.1. Methods

5.1.1. Participants

We used the same power estimate from Study 1 to determine our target sample size of $n = 750$ (150 participants in each of five conditions). We recruited a convenience sample of 802 participants from the United Kingdom through Prolific. After excluding participants who failed our pre-treatment attention check, we were left with a final sample of 753 participants (462 female; 278 male; 9 non-binary/third gender; 4 undisclosed gender; mean age = 44.29 years). 74% of these participants reported having used ChatGPT before.

5.1.2. Design

We used the same “control plus 2x2” between-subjects design as in Study 2. In the experimental conditions, we manipulated whether people in the scenarios used a standard or personalized AI model, and whether people put more or less effort into the task. This resulted in five conditions overall: (i) the control condition, (ii) the standard-low-effort

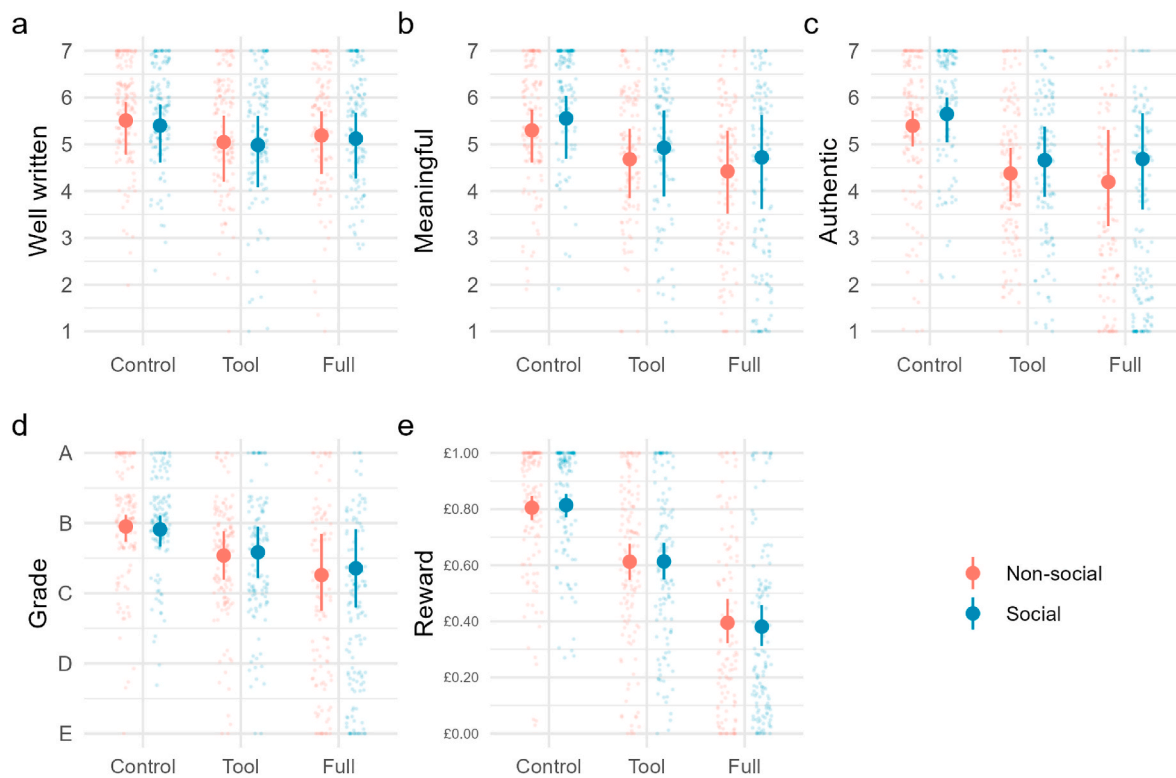


Fig. 5. Perceptions of the work in Study 3. *Note.* Participants in the control condition, the tool outsourcing condition, and the full outsourcing condition evaluated the essay response to the writing task. Participants rated whether the essay response was (a) well-written, (b) meaningful, and (c) authentic. Participants also (d) graded the work and (e) rewarded the work with a hypothetical monetary bonus. Jittered points represent participant responses to the questions, split by whether the writing task was a non-social task (red) or a social task (blue). Point ranges are estimated marginal means from the fitted model, pooling over essay answers. Points and line ranges represent posterior medians and 95% credible intervals, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

condition, (iii) the standard-high-effort condition, (iv) the personalized-low-effort condition, and (v) the personalized-high-effort condition. Our authenticity manipulation was inspired by recent psychological work looking at the credit-blame asymmetry in AI use (Earp et al., 2024), showing that people receive more personal credit for their work when they use an AI model trained on their own prior writings.

5.1.3. Procedure

The procedure was mostly identical to Study 2 to allow us to explore effects across a range of tasks, but we updated the study preamble and the presentation of the scenarios. For participants in the personalized AI conditions, we expanded the study preamble to explain that personalized AI models were trained on people's own prior writings and "tailored to each specific person and their own thoughts, feelings, and values". Then in the scenarios, we told participants in the experimental conditions:

- *Standard AI conditions:* "In order to complete this task, [the person] uses the AI tool ChatGPT."
- *Personalized AI conditions:* "In order to complete this task, [the person] uses a personalized AI tool."

We then told participants:

- *Low effort conditions:* "[The person] quickly gives the AI a rushed prompt and uses its first output."
- *High effort conditions:* "[The person] carefully gives the AI several detailed prompts and, after multiple rounds of changes, uses its resulting output."

At the end of the study, we asked participants to choose which of

these was more authentic and effortful, respectively. 94% of participants stated that the personalized AI was more authentic and 99% of participants stated that giving the AI several detailed prompts was more effortful. This suggests that even if participants might not have felt the output was meaningfully authentic in the way that mattered (see Discussion), our participants agreed that using a personalized AI was at least more authentic than using a generic one.

5.1.4. Pre-registration

We pre-registered the study on the Open Science Framework (<https://osf.io/ac9g3>).

5.1.5. Statistical analysis

We fitted the same Bayesian multivariate multilevel cumulative-link ordinal models as in Studies 1 and 2 (see Supplementary Fig. 9 for posterior predictive checks). All models converged normally ($\hat{R} \leq 1.01$; all effective sample sizes >400; no divergent transitions).

5.2. Results

We first looked across all the tasks. On average, we found that people who outsourced to AI in a low effort way were perceived as less competent, less moral, lazier, and less trustworthy than people who put more effort into their use of AI (Fig. 6; Table 5). By contrast, we found no evidence that character evaluations differed between people who used a standard AI model rather than a personalized AI model. Due to uncertainty in the parameter estimates, we were unable to draw conclusions about interaction effects between effort and AI type.

As in Studies 1 and 2, the effects of outsourcing to AI varied across the different tasks, especially for perceptions of warmth and morality

Table 4

Pairwise contrasts for perceptions of the work in Study 3.

	Response				
	Well written	Meaningful	Authentic	Grade	Reward
Effect of outsourcing type (Task type = Social)					
Tool Social - Control Social	-0.42 [-0.84 0.03]	-0.64 [-1.11 -0.08]	-1.01 [-1.47 -0.35]	-0.33 [-0.60 -0.02]	-0.20 [-0.27 -0.13]
Full Social - Control Social	-0.28 [-0.63 0.09]	-0.84 [-1.55 -0.06]	-0.95 [-1.93 -0.07]	-0.55 [-1.07 -0.04]	-0.43 [-0.50 -0.36]
Full Social - Tool Social	0.14 [-0.39 0.67]	-0.21 [-1.13 0.74]	0.04 [-1.12 1.12]	-0.23 [-0.84 0.36]	-0.23 [-0.32 -0.14]
Effect of outsourcing type (Task type = Non-social)					
Tool Non-social - Control Non-social	-0.46 [-0.92 -0.02]	-0.62 [-1.09 -0.08]	-1.03 [-1.47 -0.49]	-0.42 [-0.71 -0.07]	-0.19 [-0.26 -0.12]
Full Non-social - Control Non-social	-0.32 [-0.70 0.03]	-0.90 [-1.56 -0.08]	-1.23 [-2.09 -0.12]	-0.71 [-1.16 -0.13]	-0.41 [-0.49 -0.32]
Full Non-social - Tool Non-social	0.14 [-0.39 0.71]	-0.28 [-1.06 0.62]	-0.21 [-1.19 0.97]	-0.29 [-0.86 0.34]	-0.22 [-0.31 -0.12]
Effect of task type					
Control Social - Control Non-social	-0.10 [-0.56 0.33]	0.28 [-0.36 0.73]	0.27 [-0.21 0.64]	-0.04 [-0.27 0.17]	0.01 [-0.05 0.06]
Tool Social - Tool Non-social	-0.06 [-0.67 0.56]	0.27 [-0.61 0.99]	0.29 [-0.41 0.95]	0.04 [-0.30 0.41]	0.00 [-0.09 0.09]
Full Social - Full Non-social	-0.07 [-0.64 0.51]	0.30 [-0.59 1.15]	0.48 [-0.37 1.38]	0.09 [-0.33 0.57]	-0.02 [-0.12 0.09]
Interaction effect					
Interaction: Tool - Control	0.05 [-0.39 0.51]	-0.02 [-0.52 0.56]	0.02 [-0.51 0.62]	0.09 [-0.22 0.39]	-0.01 [-0.10 0.08]
Interaction: Full - Control	0.04 [-0.34 0.46]	0.03 [-0.59 0.73]	0.23 [-0.50 1.02]	0.13 [-0.23 0.55]	-0.03 [-0.12 0.08]
Interaction: Full - Tool	0.00 [-0.58 0.55]	0.05 [-0.74 0.89]	0.21 [-0.66 1.12]	0.05 [-0.43 0.53]	-0.01 [-0.13 0.10]

Note. Numbers reflect differences in marginal means on either a 7-point Likert scale (well-written, meaningful, authentic), a 5-point ordinal grade scale (grade), or a 0–1 sliding scale (reward). Estimates are pooled over essay answers. The bottom rows represent the interactions between outsourcing type and task type. Main numbers are posterior medians, numbers in the square brackets are 95% credible intervals.

(Fig. 7). We again found that the negative causal effects of outsourcing to AI were not uniform, but particularly strong for tasks that are social, require social skills, impact others, have important consequences, and require effort (Supplementary Figs. 10 and 11). Indeed, for tasks like writing wedding vows or writing a love letter, outsourcing to a personalized AI in a high effort way was still perceived more negatively than the control condition for all five character dimensions.

5.3. Discussion

In Study 4, we found that effort is an important mechanism by which outsourcing to AI leads to negative character evaluations. People who engaged in effortless copying of the AI's first output were perceived more negatively than people who spent time and effort crafting the AI's output with multiple prompts. Nevertheless, for social tasks like writing wedding vows or love letters, outsourcing to AI in a high effort way was still perceived more negatively than completing the task oneself.

Interestingly, we found no evidence for an effect of authenticity as proxied by the use of a personalized AI that is trained on one's own prior writings compared to a standard AI like ChatGPT. Given the uncertainty in the parameter estimates, it is difficult to draw conclusions from this pattern. It is possible that authenticity is not an important mechanism underlying the effect of outsourcing on negative character evaluations, or that it does play a small role and we were not powered enough to detect its effect. Yet, it could also suggest that even using a personalized LLM is still seen as inauthentic in the way that matters. That is, while the majority of participants in our study stated that the personalized AI was more authentic than a standard model like ChatGPT, using a personalized LLM might still not be perceived as meaningfully authentic because of its nature: that this is a feature of the phenomenon, not a bug. An AI could be perfectly trained on all apologies that a person has ever written, but one might still think that a specific apology it then generates in a new instance is not an *authentic* apology because it does not come from the kind of processes deemed important in an apology: a personal recollection of the wrong, a commitment to change. An apology written by an AI, even in someone's own style and fully "personalized", arguably is still not an actual apology.

In Study 5, we turn to look at a third potential mechanism: a perceived lack of importance attached to the task. When participants read about someone who outsources to AI in our studies, they may be inferring that they simply did not care enough about the task – "If this was important to them, they would do it themselves!". To the extent that we especially want people to care about their relationships with others –

the kind of things demonstrated through love letters, apology notes, and gift-giving – this could explain the particular negativity we see for social tasks compared to tasks like writing daily schedules, recipes, or computer code. To test this, in Study 5, we attempted to counteract inferences about care for the task by explicitly telling participants that someone had a good reason for using AI: that they really cared about the task and used AI because they wanted to get it right.

6. Study 5

6.1. Methods

6.1.1. Participants

We used the same power estimate from Study 1 to determine our target sample size of $n = 750$ (150 participants in each of five conditions). We recruited a convenience sample of 800 participants from the United Kingdom through Prolific. After excluding participants who failed our pre-treatment attention check, we were left with a final sample of 758 participants (398 female; 346 male; 8 non-binary/third gender; 6 undisclosed gender; mean age = 41.72 years). 80% of these participants reported having used ChatGPT before.

6.1.2. Design

We used the same "control plus 2x2" between-subjects design as in Studies 2 and 4. In the experimental conditions, we manipulated whether people in the scenarios used AI as a tool or "fully" outsourced to AI, and whether people had bad or good reasons for using AI. This resulted in five conditions overall: (i) the control condition, (ii) the tool-bad-reason condition, (iii) the tool-good-reason condition, (iv) the full-bad-reason condition, and (v) the full-good-reason condition.

6.1.3. Procedure

The procedure was mostly identical to Study 4, with two changes. First, we reduced the number of tasks, focusing on eight tasks (four "social" tasks and four "non-social" tasks) that fit with the manipulation of the updated design (since, for example, participants might find it difficult to see how someone could deeply value a shopping list and want to get it right). Second, we updated the presentation of the scenarios. We told participants in the experimental conditions:

- *Bad reason conditions:* "Because they are really short on time and want to complete the task quickly, [the person] uses the AI tool ChatGPT."

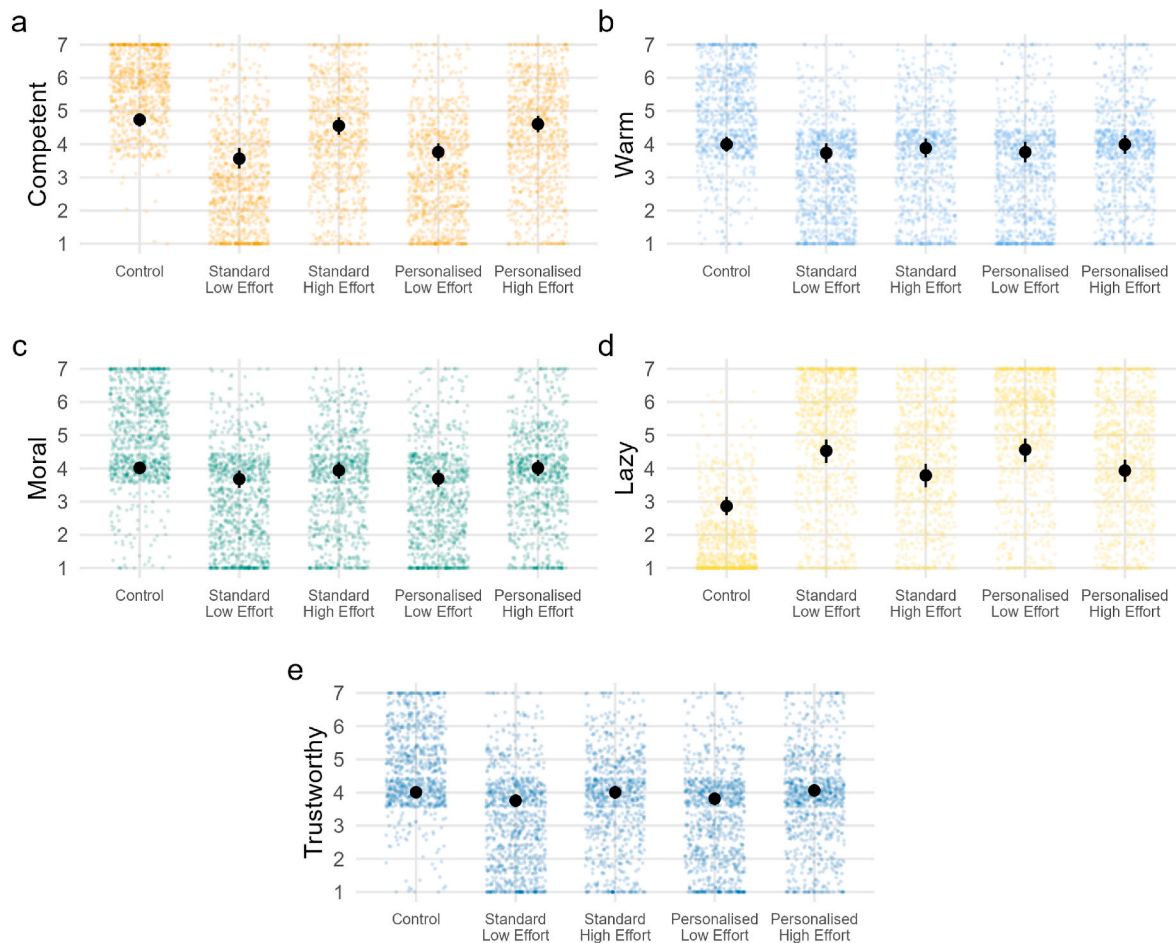


Fig. 6. Overall character evaluations in Study 4. *Note.* Participants in the control condition and four AI outsourcing conditions evaluated people in the scenarios on (a) competence, (b) warmth, (c) morality, (d) laziness, and (e) trustworthiness. Colored points represent participant responses to the questions, jittered for easier viewing. Black points are estimated marginal means from the fitted model, pooling over participants and tasks. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.

- *Good reason conditions:* “Because this task is really important to them and they want to make sure they get it right, [the person] uses the AI tool ChatGPT.”

We then told participants:

- *Tool outsourcing conditions:* “[The person] asks ChatGPT to provide ideas, inspiration, and feedback, but they edit and rewrite the suggestions and finish the task themselves.”
 - *Full outsourcing conditions:* “[The person] copies ChatGPT’s output word-for-word, rather than doing it themselves.”
- In addition to the five character evaluations, on each page we also asked participants, on a 7-point Likert scale, how much they thought the person cared about the task.

6.1.4. Pre-registration

We pre-registered the study on the Open Science Framework (<https://osf.io/ac9g3>).

6.1.5. Statistical analysis

We fitted the same Bayesian multivariate multilevel cumulative-link ordinal models as in Studies 1, 2, and 4 (see [Supplementary Fig. 12](#) for posterior predictive checks). All models converged normally ($\hat{R} \leq 1.01$; all effective sample sizes >400 ; no divergent transitions).

6.2. Results

We first looked across all the tasks. In line with our previous results, we found that people who fully outsourced to AI by copying the output verbatim were perceived as less competent, less moral, and less trustworthy than people who used AI as a collaborative tool ([Fig. 8; Table 6](#)). Perhaps surprisingly, though, people’s reasons for outsourcing to AI did not appear to influence character evaluations when pooling across all the tasks. When looking at the tasks overall, we found no evidence that character evaluations differed between people who really cared about the task and wanted to get it right and people who used AI because they were short on time and wanted to complete the task quickly. This was true both when using the AI as a tool or outsourcing in full.

Importantly, though, as in our previous studies, the type of task mattered ([Fig. 9](#)). Perceptions of outsourcing were particularly negative for tasks that are social, require social skills, impact others, have important consequences, and require effort ([Supplementary Figs. 13 and 14](#)). Indeed, for socio-relational tasks like writing an apology letter and writing wedding vows, people using AI as a tool for good reasons were still perceived more negatively than the control condition on the dimensions of warmth, morality, and laziness, though not on the dimensions of competence or trustworthiness. While our manipulation “worked” in that people who completed more instrumental, technical tasks alone were perceived as caring about the task more than when they outsourced to AI because they were in a rush, the inference that the person must care less remained for the social tasks. Even telling

Table 5

Overall pairwise contrasts in Study 4.

	Response				
	Competent	Warm	Moral	Lazy	Trustworthy
Comparison to control					
Standard Low Effort - Control	-1.18 [-1.49 -0.84]	-0.26 [-0.52 0.00]	-0.34 [-0.59 -0.10]	1.67 [1.24 2.06]	-0.26 [-0.45 -0.06]
Standard High Effort - Control	-0.19 [-0.45 0.08]	-0.11 [-0.36 0.13]	-0.08 [-0.31 0.15]	0.93 [0.54 1.31]	0.00 [-0.18 0.18]
Personalized Low Effort - Control	-0.98 [-1.25 -0.71]	-0.24 [-0.51 0.02]	-0.33 [-0.57 -0.08]	1.70 [1.28 2.09]	-0.19 [-0.38 -0.01]
Personalized High Effort - Control	-0.13 [-0.38 0.14]	0.00 [-0.24 0.23]	0.00 [-0.23 0.22]	1.08 [0.68 1.44]	0.05 [-0.12 0.23]
Effect of AI type					
Standard Low Effort - Personalized Low Effort	-0.20 [-0.54 0.17]	-0.03 [-0.36 0.30]	-0.01 [-0.32 0.30]	-0.04 [-0.51 0.44]	-0.06 [-0.29 0.17]
Standard High Effort - Personalized High Effort	-0.05 [-0.38 0.27]	-0.11 [-0.41 0.20]	-0.08 [-0.37 0.22]	-0.15 [-0.62 0.33]	-0.05 [-0.28 0.18]
Effect of effort					
Standard Low Effort - Standard High Effort	-1.00 [-1.35 -0.61]	-0.15 [-0.46 0.17]	-0.26 [-0.56 0.04]	0.74 [0.26 1.21]	-0.25 [-0.50 -0.01]
Personalized Low Effort - Personalized High Effort	-0.85 [-1.17 -0.53]	-0.24 [-0.55 0.10]	-0.32 [-0.63 -0.02]	0.63 [0.14 1.08]	-0.25 [-0.48 -0.02]
Interaction effect					
Interaction effect	-0.15 [-0.62 0.35]	0.09 [-0.36 0.52]	0.06 [-0.36 0.49]	0.10 [-0.55 0.77]	0.00 [-0.35 0.32]

Note. Numbers reflect differences in marginal means on a 7-point Likert scale, pooling over participants and tasks. The bottom row represents the interaction between AI type and effort (i.e., the difference between the differences in the rows above). Main numbers are posterior medians, numbers in the square brackets are 95% credible intervals.

participants that a person used AI as a tool to help them with more socio-relational tasks because getting the task right was really important to them still led participants think that the person cared less about that task.

Moreover, when we delved further into the task-specific estimates, we found that the reasons manipulation did indeed have an effect on character evaluations for social tasks – but not non-social tasks (Supplementary Fig. 15). When writing a bereavement card, for example, people were perceived as less warm, less moral, lazier, and less trustworthy when they used AI to save time compared to when they used it because they cared about doing the task well (but, as noted, they were still perceived as caring about those social tasks less). The same was not true for non-social tasks like writing computer code or solving a mathematical equation.

6.3. Discussion

In Study 5, we attempted to counteract the potential perception that outsourcing to AI reflects caring less about the task by explicitly informing participants about the person's reason for outsourcing: they outsourced to AI because they really cared about the task and wanted to get it right. As well as replicating our finding that fully outsourcing to AI is perceived more negatively than using AI as a tool, we also found an important effect of the reasons for outsourcing, but only for socio-relational tasks. When writing a bereavement card or an apology letter, for example, people were perceived more negatively if they used an AI tool to produce a quick output in a rush, rather than to ensure they got it right. Nonetheless, for socio-relational tasks, the “best” use of AI in this study – using AI as a tool because they cared about the task and wanted to get it right – still led to targets being perceived more negatively than if they had completed the task themselves. Indeed, for socio-relational tasks (but not the more instrumental ones), even telling participants that the person used AI because the task was important to them and they wanted to get it right was still was not enough to eliminate the perception that the person cared less about that task.

While we have so far shown varying evidence for three different mechanisms that might underlie the negative perceptions of outsourcing to AI – effort, authenticity, and caring about the task – it is likely that these mechanisms are related. For example, outsourcing to AI might indicate a lack of effort, which then might signal a lack of authenticity and reduced care in the task, leading to negative character evaluations. Moreover, the mixed effects at manipulating these variables may potentially reflect a feature of AI outsourcing, rather than a bug: using AI to complete socio-relational tasks may be perceived as more inauthentic than doing the task oneself, even if the LLM is fully personalized, and

using AI to complete socio-relational tasks may still lead to a stubborn perception that the person cares less about the task, because people infer importance and internal motivation from the degree of effort spent. Our previous studies have been unable to test causal models of second-order perceptions as we manipulated the mechanisms separately and independently. In Study 6, therefore, we bring all three mechanisms together and test their combined associations with character evaluations. To do this, we focus on a single socio-relational task – writing a love letter – which we elaborate for participants with a more detailed vignette.

7. Study 6

7.1. Methods

7.1.1. Participants

We conducted a power simulation to determine our target sample size (see Supplementary Methods). The simulation suggested that a sample size of 200 participants per condition (overall $n = 600$ for three conditions) would be required to detect a small-to-medium difference between conditions (Cohen's $d \approx 0.30$) with above 80% power.

We recruited a convenience sample of 651 participants from the United Kingdom through Prolific. After excluding participants who failed our pre-treatment attention check, we were left with a final sample of 610 participants (371 female; 233 male; 4 non-binary/third gender; 2 undisclosed gender; mean age = 42.85 years). 82% of these participants reported having used AI tools like ChatGPT before.

7.1.2. Design

We randomly allocated participants into one of three conditions in a between-subjects design: (i) the control condition, (ii) the tool outsourcing condition, or (iii) the full outsourcing condition. These conditions determined how the scenario was presented to participants.

7.1.3. Procedure

We presented participants with a vignette about a person, Adam, who is writing a love letter in a Valentine's Day card to his partner (see Supplementary Materials for full vignette wording). We told participants in each of the conditions:

- **Control condition:** “Adam decides to write the love letter in the card by himself.”
- **Tool outsourcing condition:** “Adam decides to use AI to help write the love letter in the card. He asks ChatGPT to provide ideas, inspiration, and feedback, but he edits and rewrites the suggestions and finishes writing the love letter himself.”

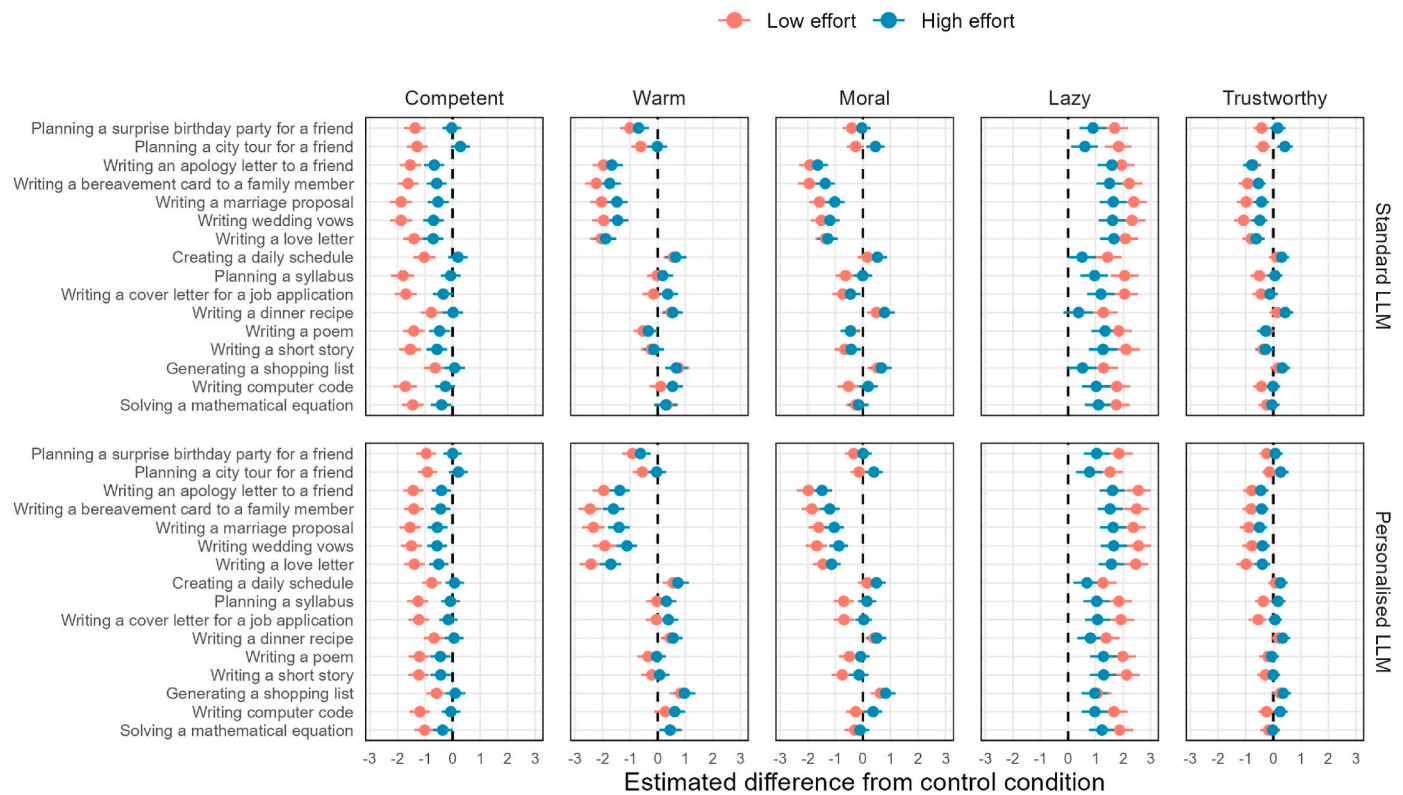


Fig. 7. Variation in the effects of outsourcing across tasks in Study 4. *Note.* Tasks are ordered from most social (top) to least social (bottom) according to ratings from a pilot study. Point ranges are differences in marginal means on a 7-point Likert scale for the low effort conditions (red) and high effort conditions (blue) compared to the control condition. Upper panels refer to the standard LLM conditions, and lower panels refer to the personalized LLM conditions. Points and ranges represent posterior medians and 95% credible intervals, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

- **Full outsourcing condition:** “Adam decides to use AI to write the love letter in the card. He asks ChatGPT to write the love letter and copies the output word-for-word, rather than writing it himself.”

We then presented participants with the love letter that Adam wrote (in reality, this was written by ChatGPT version 4o; see Supplementary Materials for wording). On the following page, we asked participants what Adam wrote and whether he used AI to help. 95% of participants answered both of these comprehension questions correctly.

Using 7-point Likert scales, we then asked participants how much effort they thought Adam put into the love letter, how authentic they thought the love letter was, how much they thought Adam cared about the love letter, and the same five character evaluations as in our previous studies. In additional free response questions, we asked participants to explain how they felt towards Adam and how they would feel if they were Adam’s partner. Finally, we asked participants several questions about AI tools like ChatGPT.

7.1.4. Pre-registration

We pre-registered the study on the Open Science Framework (<https://osf.io/ac9g3>).

7.1.5. Statistical analysis

We fitted two Bayesian regression models to the data. The first model was a multivariate cumulative-link ordinal model including all Likert scales as separate response variables. The second model was a path model capturing the effect of outsourcing on character evaluations, both directly and indirectly through perceptions of effort, authenticity, and care. In this second model, we included ordinal predictors as monotonic effects and modelled the five character evaluations as a single latent variable. We used regularizing priors for all parameters to impose

conservatism on parameter estimates (see Supplementary Methods for full model specifications and Supplementary Fig. 16 for posterior predictive checks). All models converged normally ($\hat{R} \leq 1.01$; all effective sample sizes >400 ; no divergent transitions).

7.2. Results

Across all measures, we found that outsourcing the love letter to AI was perceived more negatively compared to the control condition and that fully outsourcing to AI was perceived more negatively than using AI as a collaborative tool (Fig. 10; Table 7). Not only did outsourcing the love letter lead to more negative character evaluations, but outsourcing to AI was also seen as less effortful, less authentic, and indicative of caring less about the task.

Exploratory text analysis of participants’ free responses supported this quantitative pattern (see Supplementary Materials for methodology and Supplementary Table 8 for results). When comparing word frequencies between conditions, we found that Adam was more likely to be described as “lazy” and less likely to be described as “caring”, “thoughtful”, and “genuine” in both outsourcing conditions compared to the control condition. Adam was also more likely to be described as “romantic” and as someone who “loves” his partner when he used AI as a collaborative tool, compared to when he fully outsourced the love letter to AI.

When we included all the variables in a single path model, we found that outsourcing influenced character evaluations both directly and indirectly through our proposed mechanisms (Fig. 11). The indirect effects showed that people perceived outsourced work as less effortful, and less effortful work was seen as less authentic and indicating less care about the task. In turn, less authenticity and care were associated with more negative evaluations of the person. Effort itself was not directly

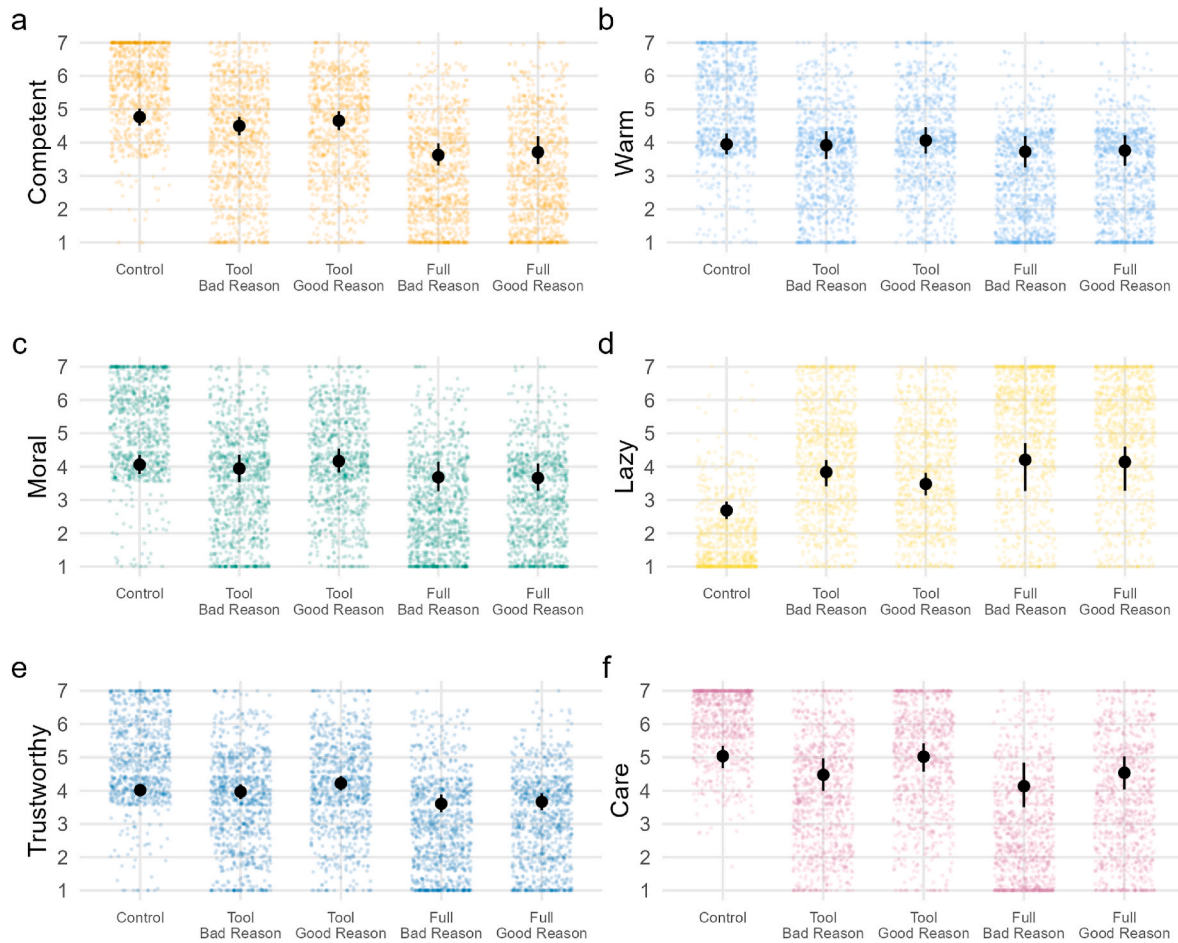


Fig. 8. Overall character evaluations in Study 5. *Note.* Participants in the control condition and four AI outsourcing conditions evaluated people in the scenarios on (a) competence, (b) warmth, (c) morality, (d) laziness, and (e) trustworthiness. Colored points represent participant responses to the questions, jittered for easier viewing. Black points are estimated marginal means from the fitted model, pooling over participants and tasks. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.

Table 6
Overall pairwise contrasts in Study 5.

	Response					
	Competent	Warm	Moral	Lazy	Trustworthy	Care
Comparison to control						
Tool Bad Reason - Control	-0.27 [-0.53 -0.02]	-0.03 [-0.36 0.30]	-0.12 [-0.46 0.23]	1.16 [0.73 1.54]	-0.05 [-0.27 0.17]	-0.56 [-0.97 -0.11]
Tool Good Reason - Control	-0.11 [-0.38 0.13]	0.11 [-0.20 0.42]	0.10 [-0.19 0.39]	0.80 [0.40 1.17]	0.21 [-0.01 0.43]	-0.02 [-0.38 0.34]
Full Bad Reason - Control	-1.15 [-1.42 -0.82]	-0.22 [-0.60 0.15]	-0.39 [-0.76 0.02]	1.57 [0.70 2.01]	-0.41 [-0.68 -0.13]	-0.92 [-1.46 -0.28]
Full Good Reason - Control	-1.07 [-1.39 -0.64]	-0.19 [-0.54 0.17]	-0.40 [-0.74 -0.05]	1.50 [0.68 1.92]	-0.35 [-0.59 -0.09]	-0.50 [-0.90 -0.05]
Effect of outsourcing type						
Full Bad Reason - Tool Bad Reason	-0.88 [-1.17 -0.52]	-0.19 [-0.68 0.30]	-0.26 [-0.76 0.25]	0.42 [-0.55 0.98]	-0.36 [-0.67 -0.04]	-0.35 [-1.04 0.38]
Full Good Reason - Tool Good Reason	-0.96 [-1.31 -0.45]	-0.31 [-0.74 0.17]	-0.50 [-0.94 -0.06]	0.70 [-0.19 1.21]	-0.56 [-0.84 -0.26]	-0.48 [-1.00 0.05]
Effect of reasons for outsourcing						
Tool Bad Reason - Tool Good Reason	-0.16 [-0.45 0.14]	-0.15 [-0.58 0.30]	-0.23 [-0.63 0.20]	0.37 [-0.15 0.84]	-0.26 [-0.54 0.02]	-0.55 [-1.04 0.03]
Full Bad Reason - Full Good Reason	-0.08 [-0.58 0.32]	-0.03 [-0.54 0.47]	0.02 [-0.49 0.53]	0.08 [-0.86 0.88]	-0.06 [-0.40 0.28]	-0.41 [-1.11 0.33]
Interaction effect						
Interaction effect	0.08 [-0.50 0.57]	0.11 [-0.58 0.77]	0.24 [-0.41 0.89]	-0.29 [-1.28 0.65]	0.19 [-0.24 0.64]	0.14 [-0.72 1.00]

Note. Numbers reflect differences in marginal means on a 7-point Likert scale, pooling over participants and tasks. The bottom row represents the interaction between outsourcing type and the reasons for outsourcing (i.e., the difference between the differences in the rows above). Main numbers are posterior medians, numbers in the square brackets are 95% credible intervals.

related to character evaluations, suggesting that effort works solely through perceptions of authenticity and care.

8. General discussion

In a world where openly available generative AI LLMs promise to let

people do more tasks, more efficiently, what is the value of doing something oneself? People can – and *do* – use AI tools like ChatGPT to, for example, create dinner recipes, assist with coding, and even write job applications (Department for Science, Innovation & Technology, 2024). But it is not only such routine, everyday, and non-social tasks that AI now “assists” with. People can use AI for a seemingly endless range of

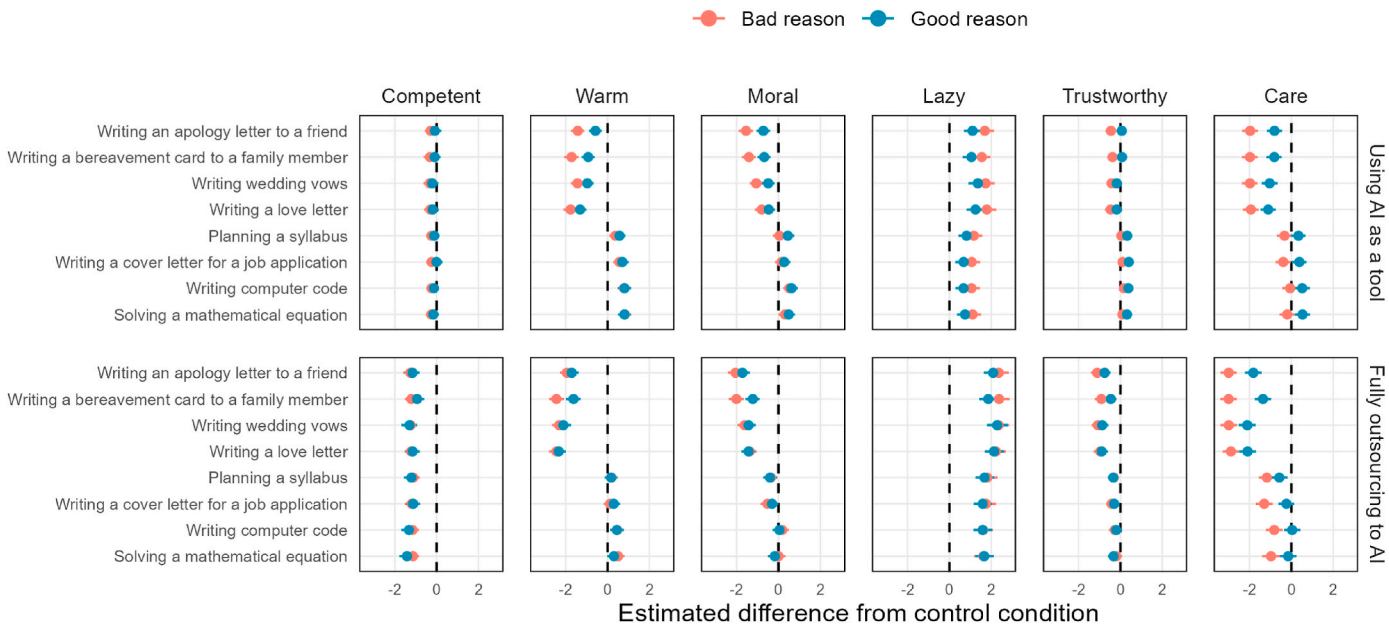


Fig. 9. Variation in the effects of outsourcing across tasks in Study 5. *Note.* Tasks are ordered from most social (top) to least social (bottom) according to ratings from a pilot study. Point ranges are differences in marginal means on a 7-point Likert scale for the bad reason conditions (red) and good reason conditions (blue) compared to the control condition. Upper panels refer to the tool outsourcing conditions, and lower panels refer to the full outsourcing conditions. Points and ranges represent posterior medians and 95% credible intervals, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

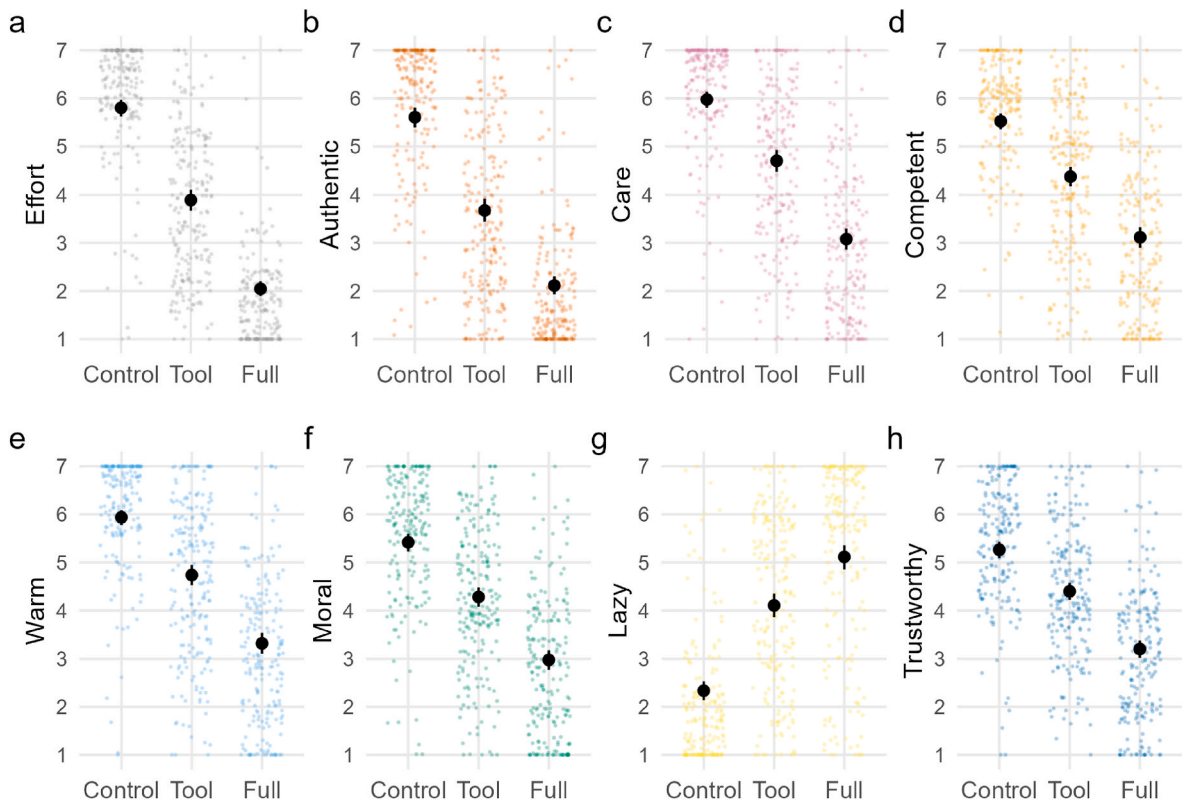


Fig. 10. Perceptions of the person and the love letter in Study 6. *Note.* Participants in the control, tool outsourcing, and full outsourcing conditions rated (a) the amount of effort put into the love letter, (b) how authentic the love letter was, (c) how much the person cared about the love letter, and (d–h) five character evaluation measures. Colored points represent participant responses to the questions, jittered for easier viewing. Black points are estimated marginal means from the fitted model. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.

Table 7
Pairwise contrasts in Study 6.

	Response							
	Effort	Authentic	Care	Competent	Warm	Moral	Lazy	Trustworthy
Tool - Control	-1.91 [-2.18 -1.64]	-1.94 [-2.23 -1.61]	-1.27 [-1.54 -1.00]	-1.15 [-1.40 -0.89]	-1.20 [-1.45 -0.95]	-1.14 [-1.39 -0.88]	1.77 [1.47 2.07]	-0.86 [-1.11 -0.62]
Full - Control	-3.76 [-3.98 -3.52]	-3.50 [-3.77 -3.22]	-2.90 [-3.16 -2.63]	-2.41 [-2.67 -2.14]	-2.62 [-2.88 -2.37]	-2.44 [-2.71 -2.17]	2.78 [2.47 3.09]	-2.07 [-2.30 -1.82]
Full - Tool	-1.85 [-2.11 -1.58]	-1.56 [-1.87 -1.26]	-1.63 [-1.93 -1.31]	-1.26 [-1.55 -0.97]	-1.42 [-1.72 -1.12]	-1.31 [-1.60 -1.02]	1.01 [0.65 1.35]	-1.20 [-1.45 -0.95]

Note. Numbers reflect differences in marginal means on a 7-point Likert scale. Main numbers are posterior medians, numbers in the square brackets are 95% credible intervals.

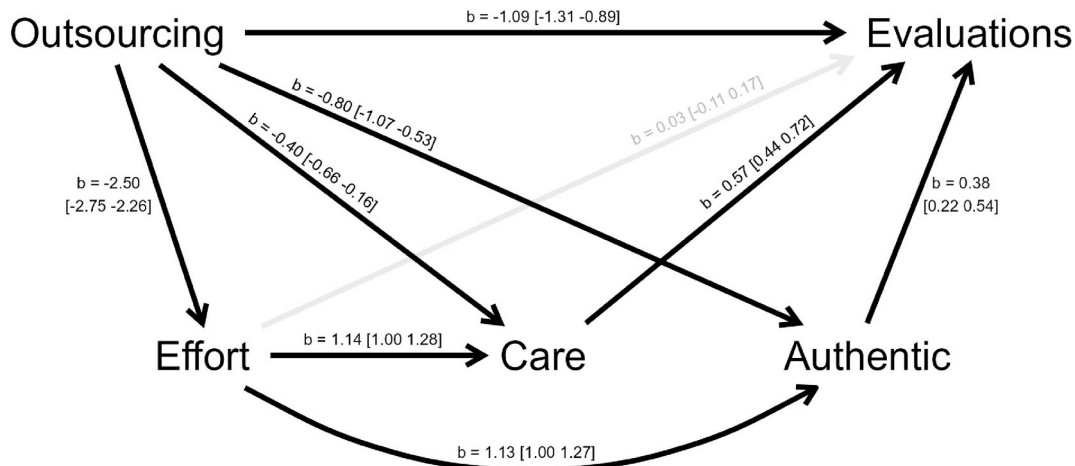


Fig. 11. Path model in Study 6. Note. All predictors were modelled as monotonic effects, such that parameters can be interpreted as the expected average difference between two adjacent categories of the ordinal predictor on the logit scale. The “evaluations” outcome variable was modelled as a single latent variable with loadings from all five character evaluations (competence, warmth, morality, laziness, and trustworthiness).

more social and consequential tasks too, from crafting apology letters to writing condolences to even writing wedding vows. But against this backdrop of increasing reliance – or trust – in AI for an increasing array of tasks, does it matter whether people do things themselves or instead “efficiently” reduce their effort – and if so, why? In this paper, across six pre-registered experiments, we show how and why AI-outsourcing shapes perceptions of others in a world where outsourcing has never been easier and cheaper.

There is an apparent paradox of effort: it is both costly and valued (see [Inzlicht et al., 2018](#)). While seemingly critical to our social lives, though, there has been relatively little attention paid to how (not) expending effort shapes how we are viewed by others. [Celniker et al. \(2023\)](#) argue that effort is moralized because people seem to infer positive moral character and cooperative intent from effort. Yet this work has looked only at a few types of cases where effort is expended or people seek outside help, where these are not necessarily the most socially impactful: looking at how people judge the quality of products like paintings, poems, or suits of armor based on how much effort is perceived ([Kruger et al., 2004](#)) or looking at how people judge others who expend more or less effort in employment and physical exercise ([Celniker et al., 2023](#)) or, in the context of AI, those who use AI for work ([Reif et al., 2025](#)). It has remained unclear how we might view others who outsource to AI specifically; how these effects might vary based on how socio-relational the task is; how different ways of outsourcing and expending less effort influence perceptions; and how outsourcing has different effects on different kinds of social perceptions. And perhaps most importantly, through asking these questions, we gain insight into the most important question: why exactly does effort have the effects that it does? Why do people care not only about the outcome achieved but the process by which it is done, and why might outsourcing to AI (or humans) trade off efficiency for inferred authenticity, care, and moral

character? Is it that expending less effort directly leads to negative social perceptions across a range of tasks because effort itself is moralized, as some have argued in psychology ([Celniker et al., 2023](#)) or, more in line with philosophical perspectives (e.g., [Aristotle, 2009](#)) might expending effort also lead to second-order judgments about how authentic someone is and how much they care about the task, which then independently drive negative perceptions? Across our studies, we shed new light on all of these questions, providing a greater understanding of how, when, and why doing something oneself the hard way shapes character judgments.

In Study 1, we showed that people who expended less effort by outsourcing tasks to AI or other humans were perceived more negatively than people who completed the tasks by themselves. At the aggregate level, these negative impressions were seen across ratings of competence, warmth, morality, laziness, and trustworthiness, but task-specific analyses revealed significant heterogeneity such that the effects were particularly strong and consistent for people who used AI to complete socio-relational tasks, such as writing a love letter or writing wedding vows. In Study 2, we focused on AI outsourcing specifically and replicated the results of Study 1, showing again that negative impressions were particularly strong for people who used AI to complete socio-relational tasks, and were found even for the “best case” of openly acknowledging the use of AI as a collaborative tool. In Study 3, we showed that people perceive both the outsourcer and the outsourced work more negatively, with outsourced work perceived as less meaningful, less authentic, and less reward-worthy than ostensibly human-generated writing. In Study 4, we showed that while it matters whether people spent time crafting the AI prompts or simply gave a rushed initial prompt, even expending effort into crafting the best prompts was still not enough to counteract the negative effects from using AI. In Study 5, we explored the potential role of the reasons for using AI and found that while explicitly telling participants that the

person used AI because they cared about the task reduced negative perceptions for social tasks, it was still not enough to eliminate negative perceptions completely. In Study 6, we showed that a perceived lack of effort is taken to signal both a lack of authenticity and lack of importance attached to the task, and these independently influenced character judgments above and beyond the effect of effort.

Together, these findings advance our understanding of how effort shapes character judgments in an AI-mediated world. We show that people negatively judge those who outsource to AI, just as we negatively judge those who outsource to humans. We show that the type of task matters: outsourcing socio-relational tasks to AI (and humans) leads to particularly negative perceptions, especially for perceptions of warmth and morality. We show that different ways of AI outsourcing lead to differences in the degree of negative perceptions but that, critically, even outsourcing to AI in the “best” way (e.g., using it as a tool and finishing the work oneself while being honest about the AI use) is still not enough to eliminate the negative consequences. And finally, advancing theoretical work on the moralization of effort (Celniker et al., 2023), we provide further insight into *why* effort matters, and why it matters in an increasingly digital age. Building on work on how expending effort is seen as a signal of moral character and cooperative intent such that effort itself becomes moralized (Celniker et al., 2023), we show that outsourcing to LLMs has consequences for perceptions of moral character. But, providing further nuance and depth to this work, we show that even if effort itself can be moralized, there is important variation across the type of tasks, the precise way that one exerts less effort, and the type of character judgments.

The first key contribution of our work is that it highlights the critical role of task type in how people are perceived from (lack of) effort. While previous work suggests that effort itself directly signals moral character and should therefore be moralized across many types of task, our results suggest that the type of task plays a critical role, and that in fact – when it comes to outsourcing to AI, at least – sometimes reduced effort does not necessarily lead to negative perceptions, at least for the traits relevant to moral character. We manipulated effort both indirectly (through describing people as either completing the task themselves; using AI as a tool for a first draft but finishing the product themselves; and fully outsourcing to AI and copying the output word-for-word) and directly (through describing the person as spending time and energy into preparing several detailed prompts vs. giving a rushed prompt and using the first output), and found consistent overall effects whereby people who expended less effort were perceived more negatively. And yet while we consistently observed overall effects of outsourcing, there was significant variation across tasks. In Study 2, for example, we found that for many of the more instrumental, non-social tasks (e.g. writing computer code, writing a short story, or planning a syllabus), those who fully outsourced to AI and copied the output word-for-word were indeed seen as less competent and more lazy, but there was no effect on moral character or perceived trustworthiness. Similar results were found in Studies 4 and 5. Our results therefore provide both support and room for nuance in work on the moralization of effort as applied to AI. In line with work on the moralization of effort (Celniker et al., 2023) expending less effort through outsourcing to AI does lead to more negative character judgments, but in contrast to what this account seems to predict, it may not always or uniformly lead to negative character judgments across all types of tasks. Perhaps in part due to the marketing and perception of AI as a tool that promises efficiency and speed, outsourcing to AI for instrumental tasks may make people seem lazier and less competent, but it may not necessarily lead to negative perceptions of their moral character. Instead, these effects on moral character are specific and pronounced for socio-relational tasks. Why? This leads us to our second key contribution.

Our second key contribution is that reduced effort leads to second order judgments about authenticity and perceived care that independently drive negative character judgments. One reason why outsourcing to AI could lead to negative character judgments is that through

outsourcing, one expends less effort, and because effort itself is moralized, this drives negative perceptions (Celniker et al., 2023). But there are also other, not mutually exclusive possibilities that are likely to be particularly important when it comes to AI use specifically. The first other possibility we considered was authenticity, drawing on findings that perceived authenticity shapes how we value objects like art (Newman & Bloom, 2012) and that perceived authenticity shapes how people interact with AI (Vo et al., 2025). If someone uses an AI tool to complete tasks for them, would they be perceived as producing work as inauthentic, and would this drive negative character judgments? Our results suggest yes. While we found no evidence that participants judged people differently if they outsourced to a completely personalized LLM vs. a generic LLM (Study 4), we did find that participant’s perceptions of how authentic AI-outsourced work would be did exert an independent effect on character judgments. The second possibility outside of effort itself that we considered was perceived care about the task. When someone chooses to outsource a task to an LLM, they may be perceived as valuing or caring about that task less, in a similar way to how people infer that people who spend more effort on a task must find it more meaningful (Campbell et al., 2025). Here, we show that people infer that when someone outsources a task to AI they must also value that task less, and this exerts an independent negative effect on character judgments outside of effort. Even directly telling participants that the task was important to them was not enough to eliminate the negative perception from outsourcing socio-relational tasks to AI. The lack of a direct effect of perceived effort in our path model showed that it is inferences of authenticity and valuing the task, rather than perceived effort per se, that are associated with negative character evaluations. As a participant in our final study put it: “*If he really cared, he would have just done it by himself from scratch*” (female, 25 years old).

Our findings that lack of effort has different effects for different tasks and that this lack of effort leads to second-order effects cohere with the philosophical idea that there is value in *how* a task was done, and not merely *whether* it was done (Aristotle, 2009; Goodman, 2010; Hursthouse & Pettigrove, 2023; Stohr, 2006). For many socio-relational tasks, it might seem that part of the constitutive action is the *process* by which it occurs: an apology that does not contain a genuine reflection and commitment to do better, rather than just the words “I am sorry”, might not seem to be an apology at all. In contrast, for many of the non-social instrumental tasks, it is easier to distinguish the importance of the process from the outcome. In this way, our work suggests that people rarely adopt a purely utilitarian perspective in which outcomes are the sole determinant (Everett & Kahane, 2020; Kahane et al., 2018). Instead, their judgments cohere more with ideas from virtue ethics about the importance of *doing* (Hursthouse & Pettigrove, 2023; Stohr, 2006). Outsourcing to AI – especially for social tasks – may allow us to produce similar outputs, but by severing the outcome from the practice of doing, it may risk the development and maintenance of our human virtues (Vallor, 2015, 2024).

AI is often being marketed as being able to help us to do more and more tasks, promising gains of efficiency that align with societal incentives for “hacks” that encourage people to do more with less energy and effort. Companies, organizations, and institutions talk about the promised benefits of AI, and the corresponding need to increase trust in these technologies. Our work, however, highlights that when it comes to our psychology, such trust has consequences and efficiency is not the only currency. Instead, *inefficiency* can sometimes pay off more, especially for social tasks. By expending effort themselves instead of outsourcing to AI, people are able to signal authenticity and care for the task, and this can lead to better reputations. Correspondingly, expending effort, even “unnecessarily”, is not as irrational, biased, or suboptimal as we might think from a utilitarian perspective in which outcomes are the only things that matter. Instead, it is precisely this “inefficiency” or “irrationality” that helps people signal things that they care about and connect with others, thereby arguably reflecting a deeply rational reflection of virtues and the importance of social ties (Everett et al.,

2016).

Most speculatively, our results on the negative effects of AI-outsourcing on character judgments highlight potential risks in how increased use of AI could lead to negative consequences for social ties, especially if people start to assume, by default, that others are using AI for the kind of tasks that matter. While our findings reflect people's social perceptions of AI use and should not be used to draw normative conclusions, other researchers have made the case that AI could have negative socio-relational effects. Sociologists have highlighted concerns about the negative effects that outsourcing to AI can have on our "connective labor", arguing that while AI can enhance certain tasks, it cannot replicate the depth of human relationships essential for effective caregiving, education, and support (Pugh, 2024). Similar arguments have been made about the risks of outsourcing empathy to AI (Landes & Everett, 2025). Trust in AI has consequences and these consequences may not be limited only to problems of people undertrusting or overtrusting AI systems. The rapid move towards using AI for more and more tasks could have serious and unintended consequences for the way we connect with one another, serving to further weaken the social ties that bind us into a community.

8.1. Limitations and directions for future research

The studies in this paper are not without their limitations. While we included a range of different socio-relational and professional tasks in an effort to improve the generalizability of our findings across domains, it would be interesting for future work to additionally explore the generalizability and variability of our findings to further tasks and over time as AI use becomes more commonplace. By focusing on generalizability across various real-world tasks in which people outsource, it could also be argued that our design lacks the richness of information in extended vignettes that might influence character evaluations.

Regarding cross-cultural generalizability, our study is limited in that we only sampled participants from the UK, a country that is relatively unique in its levels of AI development and government readiness (Oxford Insights, 2024; Tortoise Media, 2024). But attitudes towards AI vary significantly across countries. For example, a recent global report showed that trust in AI is higher in developing nations (Gillespie et al., 2025). It is therefore possible that people in these countries will have more positive perceptions of AI use, even when others use AI to complete socio-relational tasks. Future cross-cultural research should explore the generalizability of our findings to other cultures.

Finally, while we have demonstrated negative perceptions of outsourcing in this paper, it will be important for future research to explore when people might deem outsourcing to AI as acceptable or even preferable. Several of the participants in our final study expressed in their free responses that they would have been okay with Adam using AI to write the love letter if he was not a confident writer or had a learning difficulty that made writing challenging, such as dyslexia. In line with this, some research has found that people are more accepting of cognition-enhancing technologies and drugs when they are used to repair cognitive functions, rather than to enhance cognitive functions beyond "normal" levels (Medaglia et al., 2019; Rudski, 2014). Future research should explore whether negative perceptions of outsourcing persist when AI is used in a reparative way.

9. Conclusions

To conclude, across six pre-registered studies, we have demonstrated negative – but nuanced – perceptions of outsourcing to AI. We found that, overall, participants perceived individuals who outsourced tasks to AI more negatively across a range of character dimensions and perceived outsourced work as less meaningful and authentic. At the same time, however, we found that these effects were strongly sensitive to the type of task and the way that AI was used. While expending minimum effort by fully outsourcing to AI consistently led participants to perceive the

outsourcer as lazier and less competent for more instrumental, non-social tasks like planning a syllabus, writing a job application, or writing computer code, such outsourcing did not lead participants to perceive the outsourcer as less moral or trustworthy. In contrast, for socio-relational tasks like writing an apology note or love letter, there were consistent and more pronounced negative perceptions across a range of character evaluations. These negative impressions from outsourcing to socio-relational tasks were compounded when the outsourcer copied the AI's output verbatim and did not honestly acknowledge their use of AI, and held even when the outsourcer was described as putting effort into generating the prompts and when the outsourcer was described as having a good reason for using AI. Importantly, our findings show that outsourcing to AI does not only lead to negative perceptions through reduced effort, but because when people outsource socio-relational tasks to AI they are seen as being inauthentic and caring about the task less. These findings advance previous theoretical work, connect with broader debates about the importance of *doing* in social relationships, and highlight that for many tasks – especially those that are more socio-relational – it might be better to move away from a focus on making things more efficient at all costs and instead bring back a recognition of the power of *inefficiency*. Placed against the backdrop of calls to increase trust in AI to ensure widespread adoption, we show that doing something oneself, even if AI could do it quicker and easier, signals that one is authentic and cares about the task and therefore can help bind us together. In a world of algorithm-mediated interactions, AI is no substitute for investing effort into our interpersonal relationships.

CRediT authorship contribution statement

Scott Claessens: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pierce Veitch:** Writing – review & editing, Methodology, Formal analysis. **Jim A.C. Everett:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

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.

Acknowledgements

This work was generously supported by funding from the Horizon Europe UK Guarantee via the UKRI (EP/Y00440X/1) and a Philip Leverhulme Prize (PLP-2021-095) awarded to JACE. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chb.2025.108894>.

Data availability

Data and code are available at: <https://github.com/ScottClaessens/outsourcing>.

References

- Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, 128(2), 290–314. <https://doi.org/10.1037/rev0000262>

- Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(1), 1–14. <https://doi.org/10.1057/s41599-023-01787-8>
- Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., Dervieux, C., & Woodhull, G. (2024). Quarto [Computer software] Version 1.7.23. <https://doi.org/10.5281/zenodo.5960048>
- Anderson, L. (2025). How are students using ChatGPT? For therapy, breakups, and even texting friends. *Teen Vogue*. <https://www.teenvogue.com/story/how-students-using-chatgpt-therapy-breakups>
- Banks, G. C., McCauley, K. D., Gardner, W. L., & Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *The Leadership Quarterly*, 27(4), 634–652. <https://doi.org/10.1016/j.leaqua.2016.02.006>
- Campbell, A. V., Wang, Y., & Inzlicht, M. (2025). Experimental evidence that exerting effort increases meaning. *Cognition*, 257, Article 106065. <https://doi.org/10.1016/j.cognition.2025.106065>
- Celniker, J. B., Gregory, A., Koo, H. J., Piff, P. K., Ditto, P. H., & Shariff, A. F. (2023). The moralization of effort. *Journal of Experimental Psychology: General*, 152(1), 60–79. <https://doi.org/10.1037/xge0001259>
- Clark, M. S., & Mills, J. R. (2012). A theory of communal (and exchange) relationships. In P. A. M. V. Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 232–250). Sage Publications Ltd. <https://doi.org/10.4135/9781446249222.n38>
- Cubitt, R. P., Drouvelis, M., Gächter, S., & Kabin, R. (2011). Moral judgments in social dilemmas: How bad is free riding? *Journal of Public Economics*, 95(3), 253–264. <https://doi.org/10.1016/j.jpubeco.2010.10.011>
- Department for Science, Innovation & Technology. (2024). Public attitudes to data and AI: Tracker survey (Wave 4) report. <https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-4/public-attitudes-to-data-and-ai-tracker-survey-wave-4-report>
- Dutton, D. (2003). Authenticity in art. In J. Levinson (Ed.), *The Oxford handbook of aesthetics*. Oxford University Press.
- Earp, B. D., Porsdam Mann, S., Liu, P., Hannikainen, I., Khan, M. A., Chu, Y., & Savulescu, J. (2024). Credit and blame for AI-generated content: Effects of personalization in four countries. *Annals of the New York Academy of Sciences*, 1542(1), 51–57. <https://doi.org/10.1111/nyas.15258>
- Everett, J. A., & Kahane, G. (2020). Switching tracks? Towards a multidimensional model of utilitarian psychology. *Trends in Cognitive Sciences*, 24(2), 124–134. <https://doi.org/10.1016/j.tics.2019.11.012>
- Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, 145(6), 772–787. <https://doi.org/10.1037/xge0000165>
- Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99(4), 689–723. <https://doi.org/10.1037/0033-295X.99.4.689>
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Gardner, W. L., Cogliser, C. C., Davis, K. M., & Dickens, M. P. (2011). Authentic leadership: A review of the literature and research agenda. *The Leadership Quarterly*, 22(6), 1120–1145. <https://doi.org/10.1016/j.leaqua.2011.09.007>
- Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hased, G. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025*. The University of Melbourne. <https://doi.org/10.26188/28822919>. KPMG.
- Goodman, R. (2010). Cognitive enhancement, cheating, and accomplishment. *Kennedy Institute of Ethics Journal*, 20(2), 145–160. <https://doi.org/10.1353/ken.0.0309>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148–168. <https://doi.org/10.1037/a0034726>
- Heider, F. (1958). *The psychology of interpersonal relations*. John Wiley & Sons Inc. <https://doi.org/10.1037/10628-000>
- Hursthouse, R., & Pettigrove, G. (2023). Virtue ethics. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/>
- Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, 22(4), 337–349. <https://doi.org/10.1016/j.tics.2018.01.007>
- Italie, L. (2023). *What's merriam-webster's word of the year for 2023? Hint. Be true to yourself*. Associated Press News. <https://apnews.com/article/merriam-webster-word-of-year-2023-a9fea610cb32ed913bc15533acab71cc>
- Jourard, S. M. (1971). *The transparent self*. Van Nostrand Reinhold.
- Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, 125(2), 131–164. <https://doi.org/10.1037/rev0000093>
- Kernis, M. H., & Goldman, B. M. (2006). A multicomponent conceptualization of authenticity: Theory and research. *Advances in Experimental Social Psychology*, 38, 283–357. [https://doi.org/10.1016/S0065-2601\(06\)38006-9](https://doi.org/10.1016/S0065-2601(06)38006-9)
- Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, 45(4), 819–828. <https://doi.org/10.1037/0022-3514.45.4.819>
- Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology*, 40(1), 91–98. [https://doi.org/10.1016/S0022-1031\(03\)00065-9](https://doi.org/10.1016/S0022-1031(03)00065-9)
- Landau, W. M. (2021). The targets R package: A dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, 6(57), 2959. <https://doi.org/10.21105/joss.02959>
- Landes, E., & Everett, J. A. C. (2025). AI should develop human empathy, not replace it. *PsyArXiv* <https://doi.org/10.31234/osf.io/y3qzu.v1>
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Liu, B., Kang, J., & Wei, L. (2024). Artificial intelligence and perceived effort in relationship maintenance: Effects on relationship satisfaction and uncertainty. *Journal of Social and Personal Relationships*, 41(5), 1232–1252. <https://doi.org/10.1177/02654075231189899>
- Malle, B. F. (2022). Attribution theories. In *Theories in social psychology* (2nd ed., pp. 93–120). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781394266616.ch4>
- Medaglia, J. D., Yaden, D. B., Helion, C., & Haslam, M. (2019). Moral attitudes and willingness to enhance and repair cognition with brain stimulation. *Brain Stimulation*, 12(1), 44–53. <https://doi.org/10.1016/j.brs.2018.09.014>
- miramar0. (2024). My (30F) husband (36M) used AI to write his vows. How do I approach him about this? *Reddit* [Online forum post] https://www.reddit.com/r/relationship_advice/comments/1d9tw9j/my_30f_husband_36m_used_ai_to_write_his_vows_how/
- Newman, G. E. (2019). The psychology of authenticity. *Review of General Psychology*, 23(1), 8–18. <https://doi.org/10.1037/gpr0000158>
- Newman, G. E., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General*, 141(3), 558–569. <https://doi.org/10.1037/a0026035>
- Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, 22(3), 453–460. <https://doi.org/10.1016/j.jcps.2011.08.002>
- 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>
- Oxford Insights. (2024). Government AI readiness index. <https://oxfordinsights.com/ai-readiness/ai-readiness-index/>
- Pedersen, T. L. (2025). Patchwork: The composer of plots. <https://patchwork.data-imaging.inist.com>
- Pugh, A. J. (2024). *The last human job: The work of connecting in a disconnected world*. Princeton University Press.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reif, J. A., Larrick, R. P., & Soll, J. B. (2025). Evidence of a social evaluation penalty for using AI. *Proceedings of the National Academy of Sciences*, 122(19), Article e2426766122. <https://doi.org/10.1073/pnas.2426766122>
- Aristotle. (2009). *The Nicomachean ethics*. In D. Ross, & L. Brown (Eds.), *Translators*. Oxford University Press.
- Roth, L. H. O., & Tisot, T. (2025). A costly co-pilot? AI usage signals reduced morality via effort perceptions. *OSF Preprints*. <https://doi.org/10.31219/osf.io/vwxgz.v1>
- Rudski, J. M. (2014). A comparison of attitudes toward pharmacological treatment versus enhancement under competitive and noncompetitive conditions. *AJOB Empirical Bioethics*, 5(2), 80–90. <https://doi.org/10.1080/23294515.2013.850122>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68. <https://doi.org/10.1037/0003-066X.55.1.68>
- Stohr, K. (2006). Contemporary virtue ethics. *Philosophy Compass*, 1(1), 22–27. <https://doi.org/10.1111/j.1747-9991.2006.00004.x>
- Sutton, A. (2020). Living the good life: A meta-analysis of authenticity, well-being and engagement. *Personality and Individual Differences*, 153, Article 109645. <https://doi.org/10.1016/j.paid.2019.109645>
- Tait, A. (2024). My boyfriend sent me a heartfelt message - Then I learned he'd used ChatGPT. *The i Paper*. <https://inews.co.uk/inews-lifestyle/boyfriend-sent-heartfelt-message-learned-chatgpt-3413336>
- Tortoise Media. (2024). The global AI index. <https://www.tortoisemedia.com/data/global-ai>
- Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy & Technology*, 28, 107–124. <https://doi.org/10.1007/s13347-014-0156-9>
- Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in an age of machine thinking*. Oxford University Press. <https://doi.org/10.1093/oso/9780197759066.001.0001>
- Vo, D.-T., Nguyen, L. V. T., Dang-Pham, D., & Hoang, A.-P. (2025). What makes an app authentic? Determining antecedents of perceived authenticity in an AI-powered service app. *Computers in Human Behavior*, 163, Article 108495. <https://doi.org/10.1016/j.chb.2024.108495>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. <https://ggplot2.tidyverse.org>