Negative Perceptions of Outsourcing to Artificial Intelligence

Scott Claessens, Pierce Veitch, and Jim Everett

School of Psychology, University of Kent

Author Note

Scott Claessens https://orcid.org/0000-0002-3562-6981

Pierce Veitch https://orcid.org/0009-0005-3364-7470

Jim Everett https://orcid.org/0000-0003-2801-5426

All data and original code can be found on GitHub at https://github.com/ScottClaessens/outsourcing. The authors have no conflicts of interest to disclose. This work was supported by … Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Scott Claessens: data curation, formal analysis, investigation, methodology, visualization, writing, editing; Pierce Veitch: formal analysis, methodology; Jim Everett: conceptualization, funding acquisition, methodology, supervision, editing

Correspondence concerning this article should be addressed to Jim Everett, School of Psychology, University of Kent, Keynes College, Canterbury CT2 7NP, UK, Email: j.a.c.everett@kent.ac.uk

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 five pre-registered studies with British participants (n = 3649) to understand how people view those who outsource tasks to AI, and how this depends on how socio-relational the task is, whether AI is used as a tool or fully delegated to, and the acknowledgment of the AI use. We find negative perceptions of outsourcing, particularly for socio-relational tasks, and demonstrate that this does not only make us think more negatively about the outsourcer, but the outsourced work itself. We then turn to examine potential mechanisms underlying negative perceptions of outsourcing, investigating the roles of effort, perceived authenticity, and the reasons for using AI. Our research highlights the way that AI use shapes our perceptions of people, raising key philosophical questions about praise, blame, and authenticity in a world filled with AI-mediated interactions.

Keywords: artificial intelligence, character evaluation, outsourcing, trust

Negative Perceptions of Outsourcing to Artificial Intelligence

In recent years, artificial intelligence (AI) technologies have emerged as useful tools to save time and effort across a wide variety of tasks. Large language models, such as ChatGPT, Copilot, and DeepSeek, have been shown to increase efficiency for professional tasks like writing (Noy & Zhang, 2023), project planning (Barcaui & Monat, 2023), and programming (Peng et al., 2023). In a 2022 Pew survey of the American public, time-saving and efficiency were listed among the top reasons for excitement about the increased presence of AI in daily life (Rainie et al., 2022).

As outsourcing routine tasks to AI becomes more commonplace, people may begin to use AI tools to efficiently complete more socio-relational tasks. Consider, for example, writing wedding vows. While wedding vows are widely understood to be an opportunity for couples to express their love in their own words, large language models like ChatGPT will dutifully produce wedding vows upon request, without necessarily requiring any information about the partner in question. In characteristic ChatGPT style, the resulting text will likely be generic enough to apply to anyone, yet well-written enough to read aloud at the altar. A recent Reddit post from a disgruntled newlywed suggests that AI-outsourcing of this kind is already happening (miramar0, 2024).

Of course, there is nothing new about outsourcing socio-relational tasks. Ghostwriters have existed for years: in the 19th century, Victor Hugo, the writer of *Les Misérables*, used to write love letters for others to earn extra money [ref]. Today, the internet is abound with professional paid services for writing wedding vows and personal speeches. But it remains unclear how people who outsource tasks like these are perceived by others, particularly when the tasks are outsourced to impersonal AI tools like ChatGPT.

There are several reasons to expect that people will be perceived negatively for outsourcing tasks to AI. First, the effort put into a task is often used as a signal of one's moral character (Celniker et al., 2023; Tissot & Roth, 2025). People who free ride on the efforts of others are perceived as less moral (Cubitt et al., 2011), and people's outputs are perceived as more valuable the more effort was ostensibly put into them (Kruger et al., 2004). Second, people are given less credit when their work was generated jointly with or entirely by AI instead of by themselves (Earp et al., 2024). Third, from a virtue ethics perspective, people who outsource moral tasks to AI skip the opportunity to cultivate virtuous habits and develop their moral character (Vallor, 2015). These disparate lines of reasoning predict that people will be perceived more negatively if they outsource tasks to AI, though this has yet to be tested.

If people are perceived more negatively for outsourcing tasks to AI, there is also the question of how these negative perceptions might vary across different types of tasks and different kinds of AI use. Will outsourcing wedding vows to AI be perceived just as negatively as outsourcing a cover letter or a few lines of computer code? Given that effort is thought to be a particularly important cue for cooperative partner choice in social situations (Celniker et al., 2023), we predict that outsourcing socio-relational tasks to AI will be perceived more negatively than routine or professional tasks.

It is also possible for people to outsource to AI in different ways. One can complete tasks by using AI as a collaborative tool, sculpting its answer over successive prompts and editing the resulting output by hand. Alternatively, one can simply copy the AI's output verbatim. To the extent that the latter approach requires less effort, we predict that "fully" outsourcing to AI will be perceived more negatively. People can also differ in whether they are honest about their use of AI or, in the words of the disgruntled newlywed, deceptively "pass [the work] off as [their] own" (miramar0, 2024). Since people tend to dislike and distrust those who are deceptive (Tyler et al., 2006), it follows that people will perceive deceptive AI-outsourcing more negatively.

Here, we test these predictions across five pre-registered experiments with British participants. In our initial pilot studies, we found that people who outsource a range of tasks to AI are perceived more negatively than people who complete the tasks by themselves (see Supplementary Materials). In Study 1, we expand on this by manipulating the type of outsourcing to AI (as a collaborative tool vs. fully outsourcing) and whether people are honest or deceptive in their usage of AI. In Study 2, we measure perceptions of both outsourcers and the outsourced work. In Studies 3-5, we test potential mechanisms underlying negative perceptions of outsourcing, including perceived effort, perceived authenticity, and the reasons for using AI.

Study 1

Methods

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.

Participants

We conducted a power simulation to determine our target sample size. The simulation suggested that a sample size of 150 participants per condition (overall n = 750 for five 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 800 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 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 (see Supplementary Figure 1).

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.

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 16 tasks (see Supplementary Table 1 for the full list of tasks). For each scenario, we first told participants:

- *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:

- *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."

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 (see Supplementary Figure 1).

Pre-registration

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

Statistical Analysis

We fitted Bayesian multivariate multilevel cumulative-link ordinal models to the data using the *brms* R package (Bürkner, 2017). We modelled each character evaluation – competence, warmth, morality, laziness, and trustworthiness – as a separate response variable and included fixed effects for conditions, varying intercepts for participants, and varying intercepts and slopes for tasks. We used regularising priors for all parameters to impose conservatism on parameter estimates (see Supplementary Materials for full model specifications). All models converged normally ($\hat{R} \leq 1.01$).

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). Visualisations 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 data and code to reproduce the analyses and figures in this paper can be found here: https://github.com/ScottClaessens/outsourcing

Results

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 (Figure 1; Table 1). 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. We did not find any interaction effects between full outsourcing and deception.

Table 1

Pairwise Contrasts in Study 1

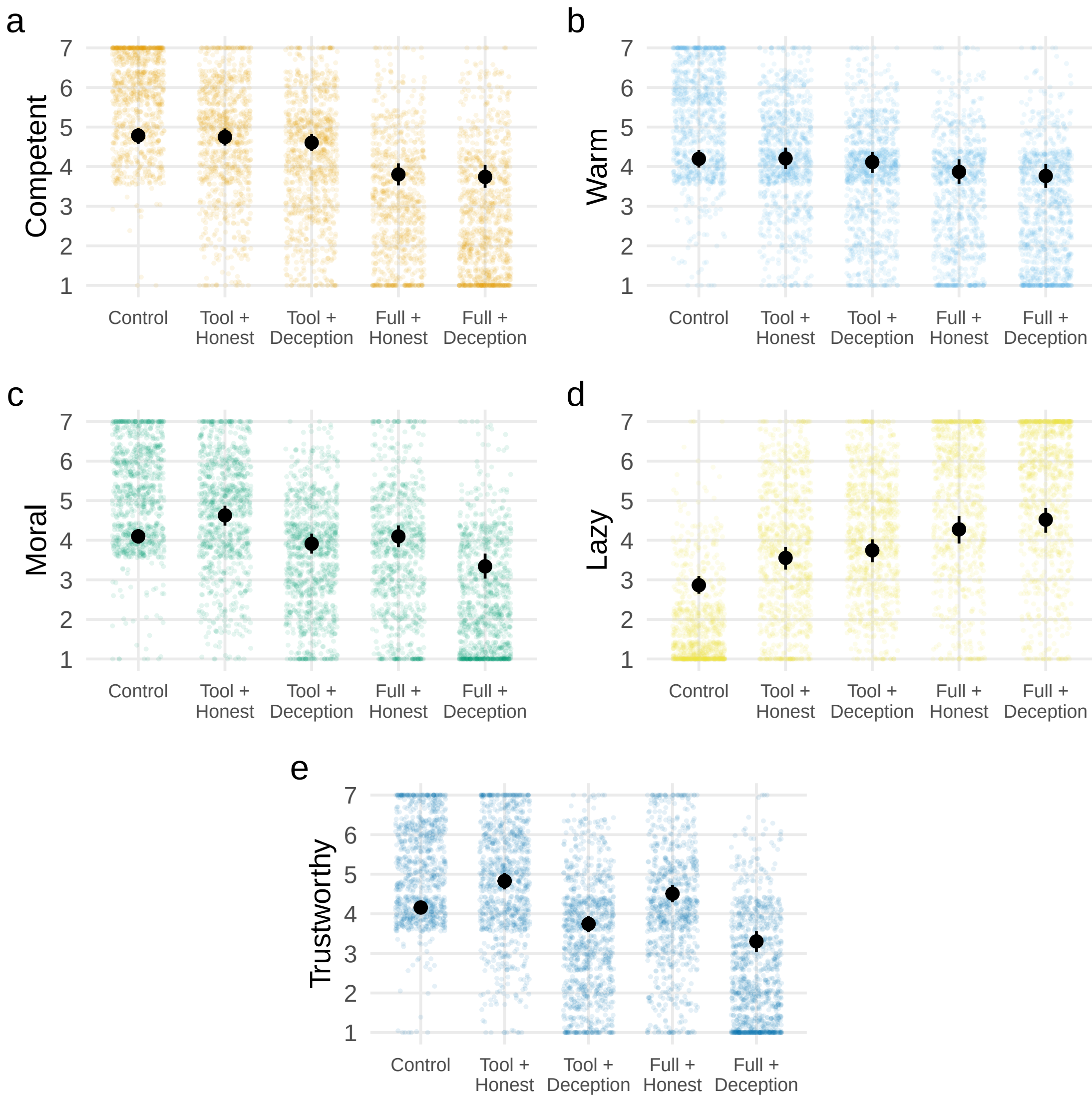
	Response				
	Competent	Warm	Moral	Lazy	Trustworthy
Comparison to control					
Tool Honest - Control	-0.04 [-0.25 0.19]	0.01 [-0.22 0.25]	0.53 [0.28 0.76]	0.69 [0.37 1.01]	0.67 [0.45 0.88]
Tool Deception - Control	-0.18 [-0.42 0.07]	-0.09 [-0.31 0.15]	-0.19 [-0.41 0.05]	0.88 [0.55 1.20]	-0.42 [-0.62 -0.20]
Full Honest - Control	-0.98 [-1.25 -0.71]	-0.33 [-0.59 -0.07]	-0.01 [-0.27 0.27]	1.42 [1.03 1.77]	0.35 [0.13 0.58]
Full Deception - Control	-1.05 [-1.33 -0.75]	-0.43 [-0.69 -0.16]	-0.76 [-1.04 -0.47]	1.66 [1.28 2.00]	-0.86 [-1.11 -0.59]
Effect of full outsourcing					
Full Honest - Tool Honest	-0.94 [-1.26 -0.63]	-0.34 [-0.64 -0.03]	-0.53 [-0.87 -0.20]	0.72 [0.29 1.14]	-0.32 [-0.60 -0.03]
Full Deception - Tool Deception	-0.87 [-1.18 -0.52]	-0.36 [-0.66 -0.02]	-0.57 [-0.91 -0.24]	0.78 [0.35 1.17]	-0.44 [-0.73 -0.15]
Effect of deception					
Tool Deception - Tool Honest	-0.14 [-0.42 0.13]	-0.09 [-0.39 0.20]	-0.72 [-1.01 -0.40]	0.19 [-0.19 0.58]	-1.08 [-1.35 -0.81]
Full Deception - Full Honest	-0.06 [-0.41 0.31]	-0.11 [-0.43 0.24]	-0.76 [-1.12 -0.38]	0.24 [-0.21 0.69]	-1.21 [-1.52 -0.89]
Interaction effect					
Interaction effect	0.08 [-0.36 0.53]	-0.01 [-0.45 0.43]	-0.03 [-0.51 0.42]	0.06 [-0.55 0.64]	-0.12 [-0.53 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 effects of outsourcing to AI varied across the different tasks, especially for perceptions of warmth and morality (Figure 2). 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 regardless of whether AI was used as a tool or fully delegated to and whether the person was honest or deceptive about their use of AI. By contrast, we observed weaker effects of outsourcing for non-social tasks like writing computer code or planning a syllabus.

Figure 1

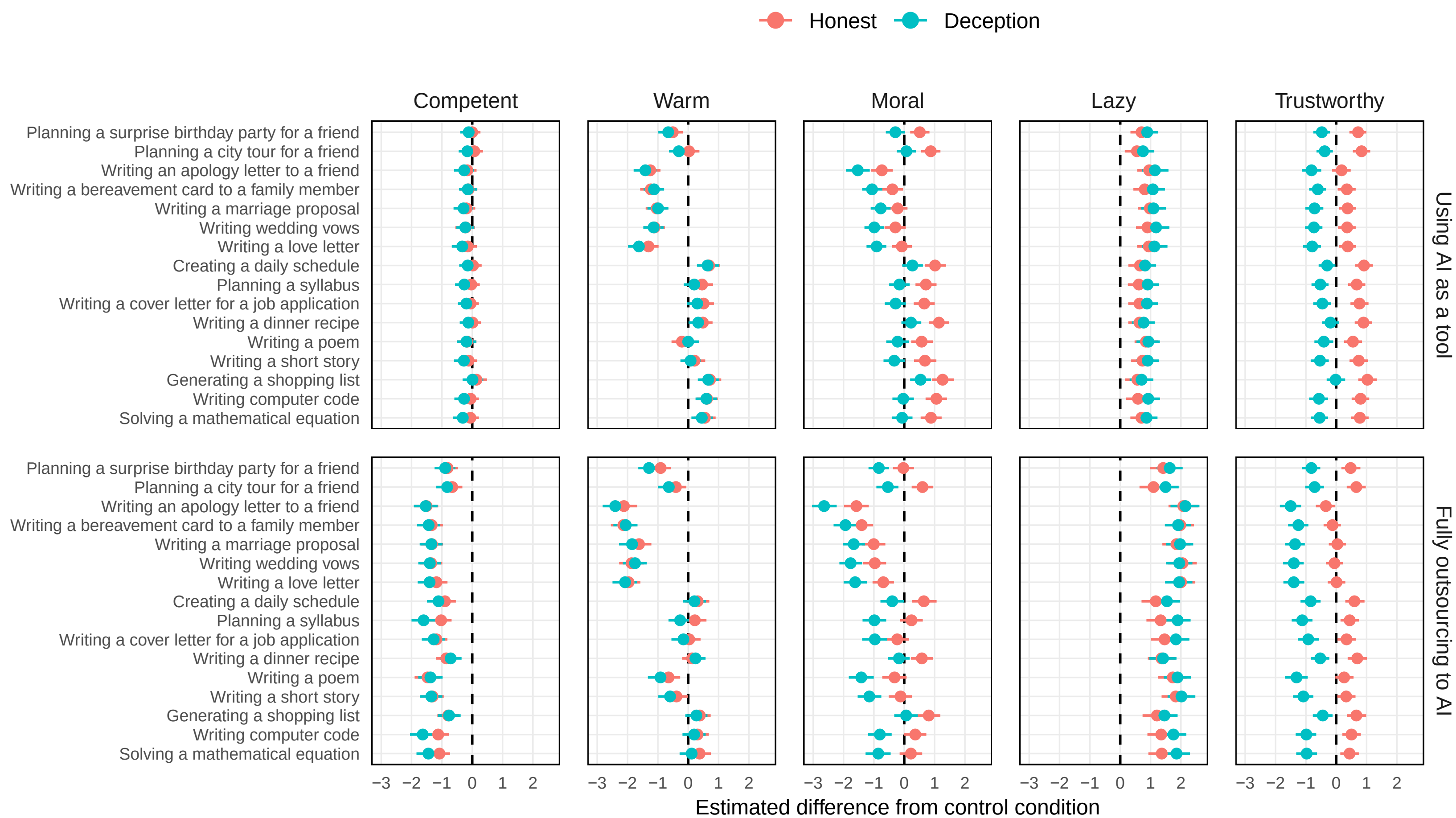
Character Evaluations in Study 1



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. Coloured 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.

Figure 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 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.

To determine the factors that predict variation across tasks, we incorporated ratings of tasks from a pilot study (see Supplementary Materials for details). Participants were asked to rate the different 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 to AI compared to the control (Supplementary Figures 2 and 3). In other words, outsourcing to AI is perceived more negatively for tasks that have these features, compared to tasks without these features.

Discussion

In this study, we manipulated whether people used AI as a collaborative tool or copied the AI's output verbatim. In line with our predictions, we found that "fully" outsourcing to AI was perceived more negatively than using AI as a collaborative tool, particularly for socio-relational tasks. These negative perceptions of people who fully outsource their work to AI may be driven in part by a general suspicion in unsupervised AI output due to what has been termed "algorithmic aversion" – the tendency for people to distrust AI relative to humans (Dietvorst et al., 2015). This aversion is particularly strong in the moral domain (Bigman & Gray, 2018). Our results suggest that people may not only harbour negative perceptions of AIs themselves, but also of humans who unthinkingly copy the output from these AIs, especially for socio-relational tasks.

We also manipulated whether people were honest or deceptive in their use of AI. We found that people were seen as less moral and less trustworthy if they did not acknowledge their use of AI when asked how they came up with their ideas. This builds on previous work showing that people dislike and distrust those who lie (Tyler et al., 2006). However, passing off AI-generated work as one's own is a specific kind of lie – it involves taking undeserved credit for the work of a machine and misrepresenting one's own emotional capabilities and writing skills. Interestingly, we found that failing to acknowledge use of AI had negative effects on perceptions of morality and trustworthiness across all the tasks we studied, suggesting that this phenomenon generalises across a range of socio-relational and professional tasks.

We have shown outsourcing to AI leads to negative character evaluations, especially when the AI output is copied verbatim and the work is passed off as one's own. However, it remains unclear whether these negative perceptions also extend to the work itself. It could be, for example, that someone is perceived badly for using ChatGPT to write their bereavement card, but the actual quality of the writing is perceived to be just as good, if not better, 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). However, because this previous study used treatment-blind raters, it was unable to assess whether knowing that a piece of text was AI-generated affects how people perceive the text, while holding the quality of the text constant. We explored this possibility in Study 2.

Study 2

Methods

Participants

We conducted a power simulation to determine our target sample size. 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 (see Supplementary Figure 4).

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. We also manipulated whether the task prompt was social or non-social.

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 and edited the text to appear more human-like). 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. 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."

We then presented participants with a randomly-chosen pre-generated essay answer to the prompt (see Supplementary Tables 2 and 3 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 4).

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. Responses to these questions suggested that the outsourcing manipulation was successful (see Supplementary Table 5). We also asked participants several questions about ChatGPT (see Supplementary Figure 4).

Pre-registration

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

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 regularising priors for all parameters to impose conservatism on parameter estimates (see Supplementary Materials for full model specifications). All models converged normally ($\hat{R} \leq 1.01$).

[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.

Results

We found that essay responses that were ostensibly generated using AI were perceived as less meaningful and less authentic, but just as well-written, compared to essay responses ostensibly written by a human (Figure 3; Table 2). Participants also marked AI-generated essays with a lower grade and rewarded AI-generated essays with a lower hypothetical monetary bonus.

Table 2

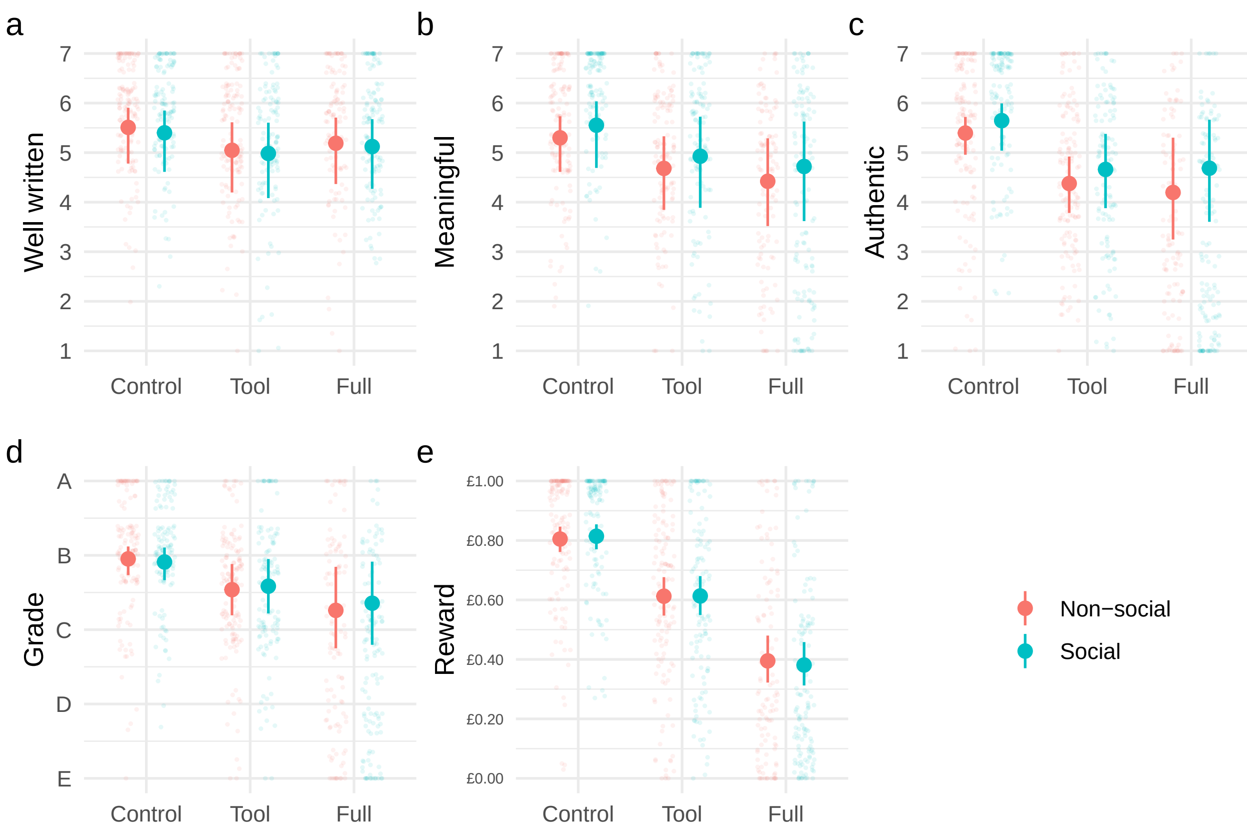
Pairwise Contrasts for Perceptions of the Work in Study 2

	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]
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.

Figure 3

Perceptions of the Work in Study 2



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.

We did not find differences in perceptions of the work between 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. We did not find any differences between social and non-social tasks and did not find any interaction effects.

Turning to character evaluations, we found that people were 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 Figure 5; Supplementary Table 6). However, we did not find any differences in character evaluations between the tool outsourcing and full outsourcing conditions. We did not find differences in character evaluations between social and non-social tasks and did not find any interaction effects.

Discussion

Extending our previous results, we found that not only are people perceived negatively for outsourcing their work to AI, but the work produced by AI is itself also perceived negatively. In particular, text that has purportedly been generated using AI is perceived to be less meaningful, less authentic, worthy of a lower grade, and worthy of a lower monetary reward compared to human-generated text. This is despite the actual text itself being held constant, perhaps explaining why perceived writing quality did not consistently differ between conditions.

Unlike in Study 1, we found no differences in the effect of AI-outsourcing between social and non-social tasks. This may be due to the particular tasks we chose. Writing *about* someone close to you is not quite the same as writing something *for* someone close to you, as is the case with wedding vows, love letters, and bereavement cards. We also found no differences between the tool and full outsourcing conditions, aside from the lower rewards given to participants in the latter condition. It is possible that differences between conditions were smaller under this design and we were not sufficiently powered to detect them.

Thus far, we have presented evidence that outsourcing tasks to AI causes negative character evaluations and negative perceptions of the work. But the mechanisms underlying these effects remain unclear. We have already suggested that effort may play a role, since perceived effort (or lack thereof) is often used as a signal of one's moral character (Cubitt et al., 2011) and cooperative intent (Celniker et al., 2023). Study 2 also suggested a role of authenticity: people who outsource to AI may be perceived as producing work that is less authentically their own, leading to negative evaluations (see also Kirk & Givi, 2025). Indeed, previous studies have shown that people react negatively to individuals and brands that behave inauthentically by strategically presenting themselves in ways that are misaligned with their true nature (Sedikides & Schlegel, 2024; Silver et al., 2021).

To explore the potential mechanisms of effort and authenticity, we return to our previous design and experimentally manipulate (1) how much effort people put into the task and (2) whether people outsource the task to a standard large language model like ChatGPT or a personalised model that has been trained specifically on one's own prior writings. The latter authenticity manipulation is inspired by recent psychological work on the credit-blame asymmetry in AI use showing that people receive more personal credit for their work when they use an AI model trained on their own prior writings (Earp et al., 2024). If effort and authenticity are important mechanisms underlying negative perceptions of outsourcing to AI, then these manipulations should influence character evaluations.

Study 3

Methods

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 (see Supplementary Figure 6).

Design

We used the same "control plus 2x2" between-subjects design as in Study 1. In the experimental conditions, we manipulated whether people in the scenarios used a standard or personalised 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 condition, (*iii*) the standard-high-effort condition, (*iv*) the personalised-low-effort condition, and (*v*) the personalised-high-effort condition.

Procedure

The procedure was mostly identical to Study 1, but we updated the study preamble and the presentation of the scenarios. For participants in the personalised AI conditions, we expanded the study preamble to explain that personalised 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."
- *Personalised AI conditions*: "In order to complete this task, [the person] uses a personalised 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 personalised AI was more authentic and 99% of participants stated that giving the AI several detailed prompts was more effortful.

Pre-registration

We pre-registered the study on the Open Science Framework (https://osf.io/vaq7u).

Statistical Analysis

We fitted the same Bayesian multivariate multilevel cumulative-link ordinal models as in Studies 1 and 2. All models converged normally ($\hat{R} \leq 1.01$).

Results

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 (Figure 4; Table 3). By contrast, we found that character evaluations did not differ between people who used a standard AI model rather than a personalised AI model. We also found no interaction effects between effort and the type of AI used.

As in Study 1, the effects of outsourcing to AI varied across the different tasks, especially for perceptions of warmth and morality (Supplementary Figure 7). We again found that the negative causal effects of outsourcing to AI were particularly strong for tasks that are social, require social skills, impact others, have important consequences, and require effort (Supplementary Figures 8 and 9).

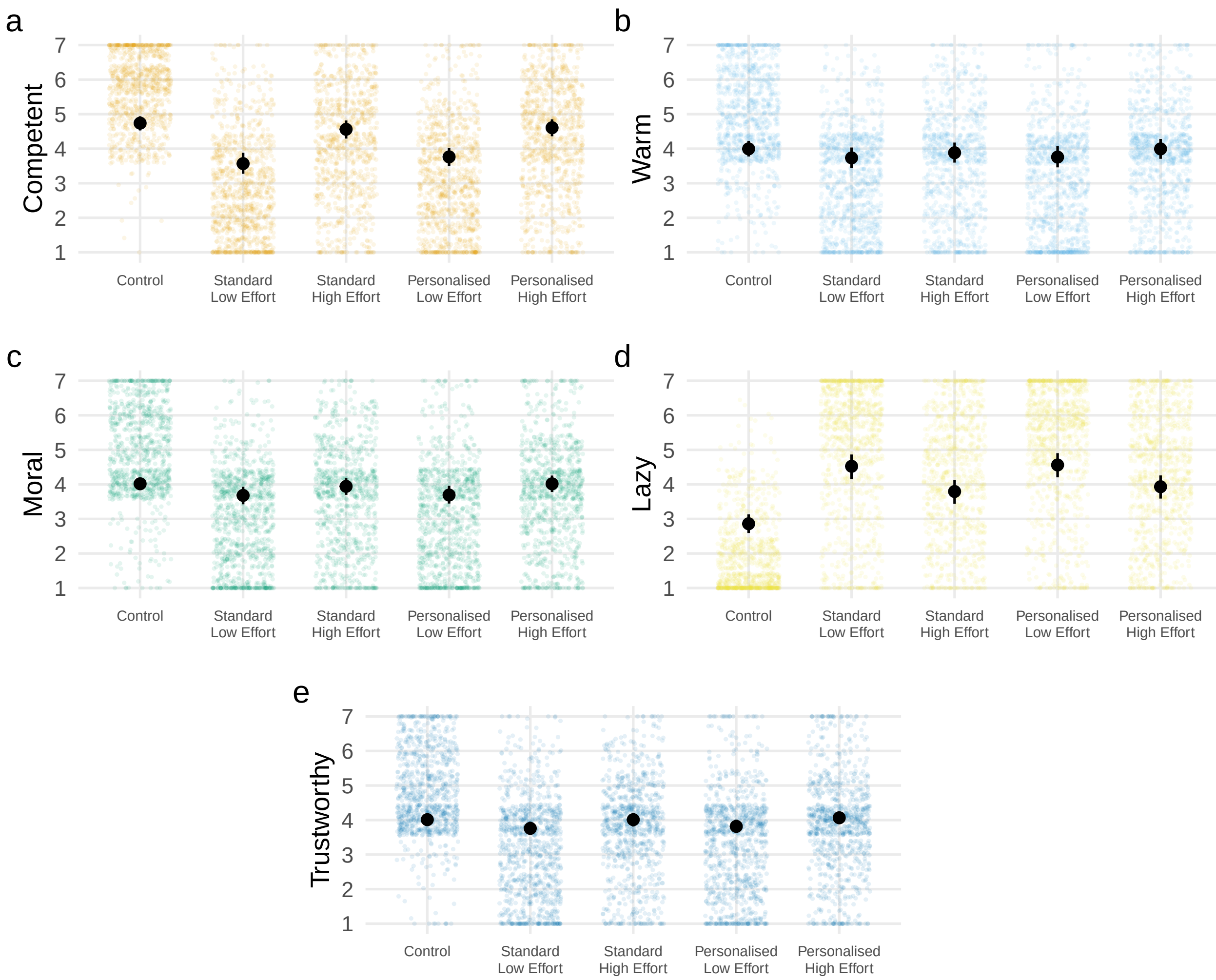
Discussion

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, whereas people who put in effort with multiple detailed prompts were perceived no differently from those who completed the tasks themselves. This result is in line with previous work showing that effort is a signal of one's moral character (Celniker et al., 2023; Cubitt et al., 2011).

Interestingly, we found no effect of authenticity as proxied by the use of a personalised AI that is trained on one's own prior writings compared to a standard AI like ChatGPT. This perhaps indicates that authenticity is not an important mechanism underlying the effect of outsourcing on

Figure 4

Character Evaluations in Study 3



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. Coloured 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 3

Pairwise Contrasts in Study 3

	Response				
	Competent	Warm	Moral	Lazy	Trustworthy
Comparison to control					
Standard Low Effort - Control	-1.17 [-1.49 -0.85]	-0.26 [-0.53 0.00]	-0.34 [-0.58 -0.10]	1.67 [1.26 2.05]	-0.25 [-0.44 -0.06]
Standard High Effort - Control	-0.18 [-0.44 0.09]	-0.11 [-0.35 0.12]	-0.08 [-0.31 0.14]	0.94 [0.54 1.32]	0.00 [-0.19 0.19]
Personalised Low Effort - Control	-0.98 [-1.24 -0.70]	-0.24 [-0.50 0.02]	-0.33 [-0.56 -0.08]	1.71 [1.29 2.08]	-0.19 [-0.38 -0.01]
Personalised High Effort - Control	-0.13 [-0.38 0.13]	0.00 [-0.24 0.23]	0.00 [-0.22 0.23]	1.08 [0.69 1.44]	0.05 [-0.12 0.23]
Effect of AI type					
Standard Low Effort - Personalised Low Effort	-0.19 [-0.54 0.17]	-0.03 [-0.35 0.31]	-0.01 [-0.32 0.30]	-0.04 [-0.51 0.43]	-0.06 [-0.30 0.18]
Standard High Effort - Personalised High Effort	-0.04 [-0.36 0.27]	-0.11 [-0.41 0.20]	-0.07 [-0.36 0.22]	-0.14 [-0.61 0.34]	-0.05 [-0.28 0.17]
Effect of effort					
Standard Low Effort - Standard High Effort	-1.00 [-1.34 -0.62]	-0.15 [-0.47 0.17]	-0.26 [-0.56 0.04]	0.73 [0.25 1.21]	-0.25 [-0.50 -0.02]
Personalised Low Effort - Personalised High Effort	-0.84 [-1.16 -0.53]	-0.24 [-0.56 0.10]	-0.33 [-0.63 -0.03]	0.63 [0.15 1.08]	-0.25 [-0.48 -0.03]
Interaction effect					
Interaction effect	-0.15 [-0.61 0.32]	0.09 [-0.36 0.53]	0.06 [-0.36 0.49]	0.10 [-0.55 0.76]	0.00 [-0.32 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.

negative character evaluations. However, our specific manipulation may not have moved the needle on authenticity enough to impact character evaluations. While previous work has found an effect of personalised AI models on perceived credit (Earp et al., 2024), and the majority of participants in our study stated that the personalised AI was more authentic than a standard model like ChatGPT, it is possible that perceptions of authenticity in our study remained low even with the personalised AI model.

Beyond effort and authenticity, another mechanism that may be driving negative perceptions of outsourcing to AI is people's reasons for outsourcing. People may use AI for bad reasons, such as saving time or strategically presenting oneself differently. But there are also good reasons for using AI. For example, someone might use AI because they really care about the task and want to make sure they get it right. If communicated to participants, will these reasons influence perceptions of outsourcers? We answer this question in Study 4.

Study 4

Methods

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 (see Supplementary Figure 10).

Design

We used the same "control plus 2x2" between-subjects design as in Studies 1 and 3. 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.

Procedure

The procedure was mostly identical to Study 3, with two changes. First, we reduced the number of tasks, focusing on eight tasks (four "social" tasks and four "non-social" tasks) that made sense in our updated design and showed the largest effects in our previous studies. 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."

- *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.

Pre-registration

We pre-registered the study on the Open Science Framework (https://osf.io/vaq7u).

Statistical Analysis

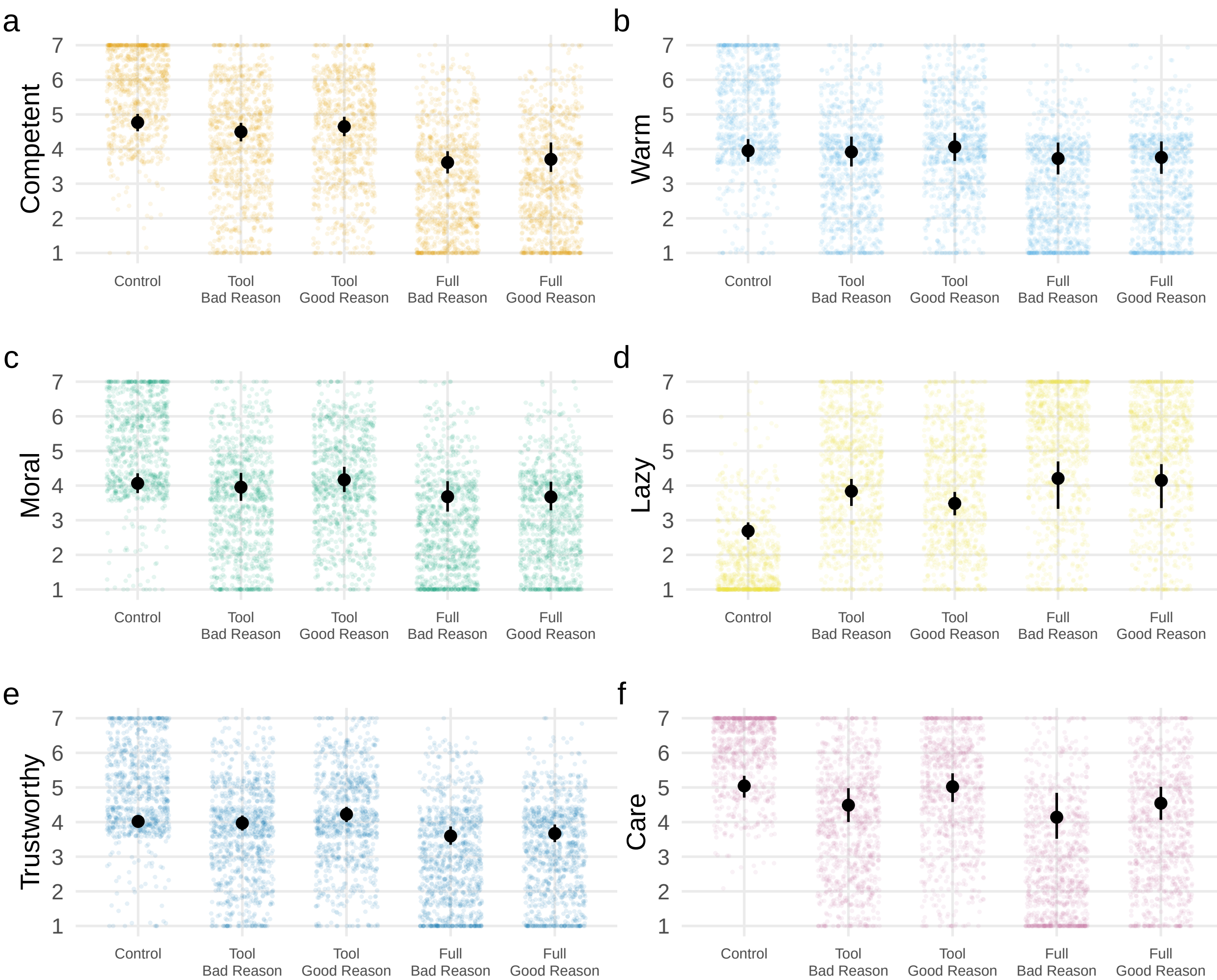
We fitted the same Bayesian multivariate multilevel cumulative-link ordinal models as in Studies 1 and 3. All models converged normally ($\hat{R} \leq 1.01$).

Results

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 (Figure 5; Table 4). By contrast, people's reasons for outsourcing to AI did not appear to influence character evaluations when pooling across all the tasks. We also found no interaction effects between the type of outsourcing and the reasons for outsourcing.

Figure 5

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. Coloured 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 4

Pairwise Contrasts in Study 4

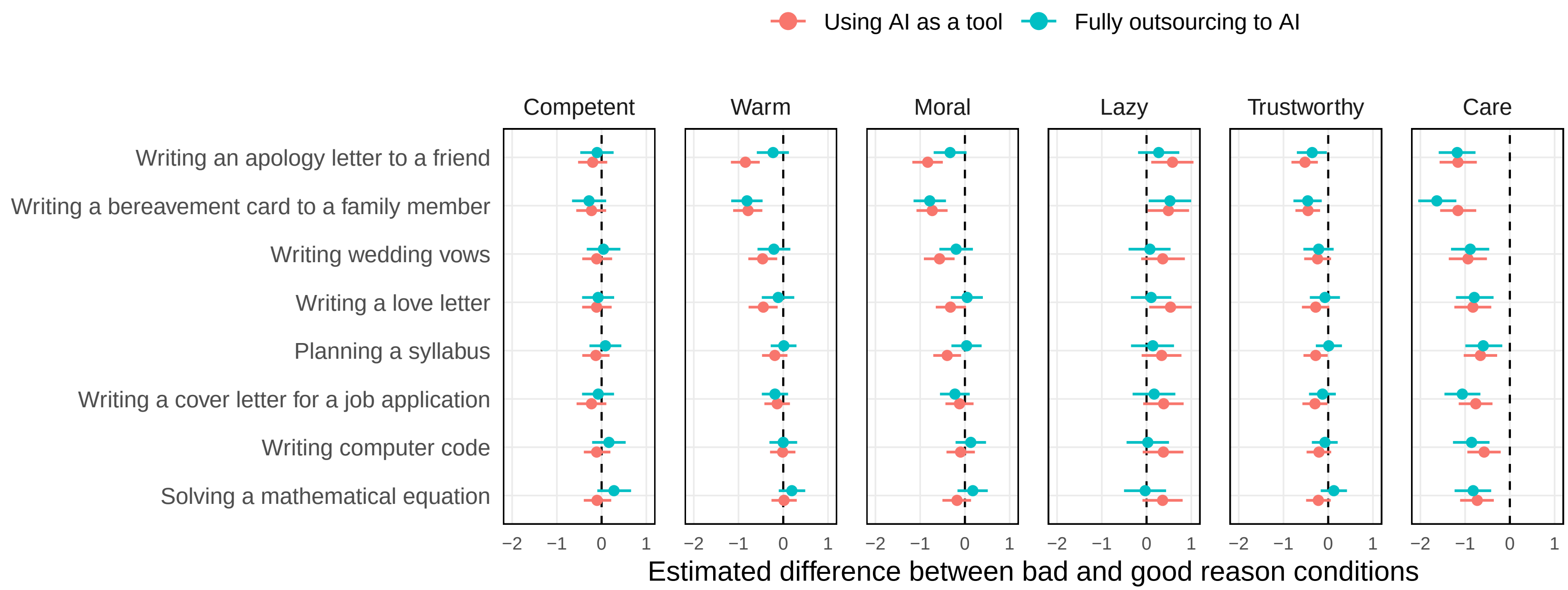
	Response					
	Competent	Warm	Moral	Lazy	Trustworthy	Care
Comparison to control						
Tool Bad Reason - Control	-0.27 [-0.53 -0.02]	-0.03 [-0.37 0.30]	-0.11 [-0.45 0.24]	1.16 [0.73 1.51]	-0.05 [-0.27 0.16]	-0.56 [-0.97 -0.11]
Tool Good Reason - Control	-0.12 [-0.38 0.13]	0.11 [-0.21 0.43]	0.11 [-0.19 0.39]	0.80 [0.43 1.17]	0.21 [-0.02 0.42]	-0.02 [-0.37 0.31]
Full Bad Reason - Control	-1.16 [-1.45 -0.83]	-0.22 [-0.59 0.14]	-0.39 [-0.75 0.00]	1.58 [0.75 2.01]	-0.42 [-0.68 -0.15]	-0.92 [-1.46 -0.29]
Full Good Reason - Control	-1.08 [-1.40 -0.62]	-0.19 [-0.54 0.16]	-0.40 [-0.73 -0.05]	1.50 [0.78 1.94]	-0.35 [-0.59 -0.09]	-0.50 [-0.91 -0.07]
Effect of outsourcing type						
Full Bad Reason - Tool Bad Reason	-0.89 [-1.19 -0.55]	-0.19 [-0.66 0.27]	-0.27 [-0.77 0.21]	0.42 [-0.53 0.98]	-0.37 [-0.68 -0.05]	-0.36 [-1.04 0.39]
Full Good Reason - Tool Good Reason	-0.96 [-1.32 -0.46]	-0.30 [-0.78 0.16]	-0.50 [-0.93 -0.04]	0.70 [-0.11 1.24]	-0.56 [-0.85 -0.23]	-0.48 [-0.99 0.07]
Effect of reasons for outsourcing						
Tool Bad Reason - Tool Good Reason	-0.15 [-0.44 0.14]	-0.14 [-0.59 0.28]	-0.22 [-0.63 0.23]	0.36 [-0.16 0.82]	-0.25 [-0.53 0.03]	-0.54 [-1.04 0.01]
Full Bad Reason - Full Good Reason	-0.07 [-0.60 0.31]	-0.03 [-0.54 0.45]	0.01 [-0.48 0.49]	0.08 [-0.85 0.85]	-0.07 [-0.41 0.26]	-0.42 [-1.10 0.34]
Interaction effect						
Interaction effect	0.07 [-0.50 0.55]	0.11 [-0.57 0.75]	0.22 [-0.41 0.86]	-0.28 [-1.28 0.64]	0.18 [-0.26 0.61]	0.13 [-0.73 1.01]

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.

As in our previous studies, we found that the effects of outsourcing to AI varied across the different tasks (Supplementary Figure 11) and were particularly negative for tasks that are social, require social skills, impact others, have important consequences, and require effort (Supplementary Figures 12 and 13). Moreover, the task-specific estimates revealed that the reasons manipulation had an effect on character evaluations for social tasks, rather than non-social tasks (Figure 6). 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. The same was not true for non-social tasks like writing computer code or solving a mathematical equation.

Figure 6

Variation in the Effect of Reasons 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 between the "bad reason" and "good reason" conditions, split by outsourcing type. Points and ranges represent posterior medians and 95% credible intervals, respectively.

Discussion

In Study 4, we replicated our previous result that fully outsourcing to AI is generally perceived more negatively than using AI as a collaborative tool. We also found that people's reasons for using AI influenced character evaluations, but only for social tasks. For example, when writing a bereavement card or an apology letter, 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. When writing a cover letter or solving a mathematical equation, we found no effects of the reasons manipulation.

Why do the reasons for using AI matter specifically for socio-relational tasks? One possibility is again to do with effort as a signal of relationship commitment. For writing computer code, it does not matter if someone uses AI to find a quicker, more efficient way of working. But for socio-relational tasks, using AI to cut corners signals a lack of commitment to and care about the relationship. In our design, we explicitly told participants about the person's motivations for using AI, so participants no longer needed to infer lack of care from the signal of reduced effort.

To sum up so far, we have suggested and found varying evidence for three different mechanisms that might underlie the negative perceptions of outsourcing to AI: effort, authenticity, and caring about the task. However, 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. Our previous studies have been unable to test causal models like these as we manipulated the mechanisms separately and independently.

In our final study, we bring all three mechanisms together and test their combined associations with character evaluations. Although we previously found no effect of authenticity as proxied by the type of AI used, it is possible that this manipulation was not a good proxy, and so we reevaluate authenticity here. We also focus on a single socio-relational task – writing a love letter – which we elaborate for participants with a more detailed vignette.

Study 5

Methods

Participants

We conducted a power simulation to determine our target sample size. 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 (see Supplementary Figure 14).

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.

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."
- *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 (see Supplementary Figure 14).

Pre-registration

We pre-registered the study on the Open Science Framework (https://osf.io/k9v7z).

Statistical Analysis

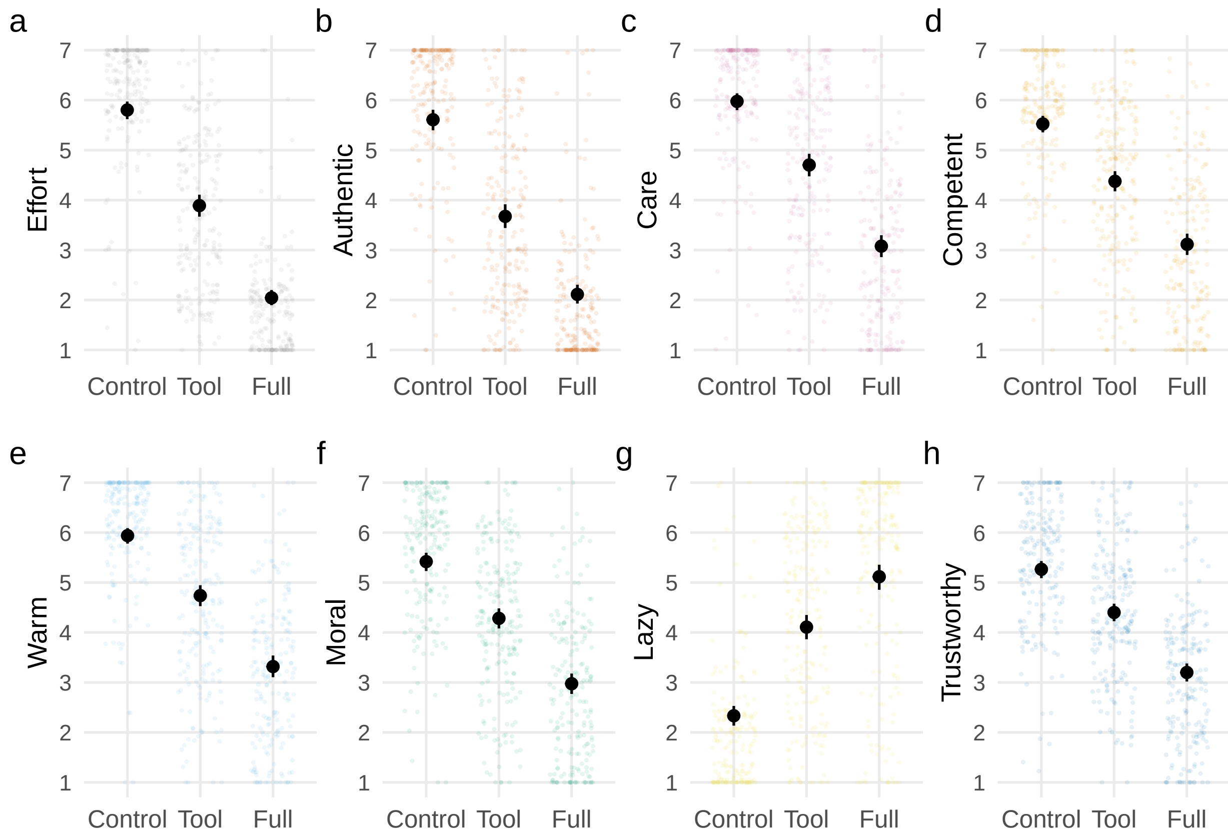
We fitted two Bayesian regression models to the data. The first model was a multivariate cumulative-link ordinal model including all Likert scales as separate response variables. The second model was a path model capturing the effect of outsourcing on character evaluations, both directly and indirectly 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 regularising priors for all parameters to impose conservatism on parameter estimates (see Supplementary Materials for full model specifications). All models converged normally ($\hat{R} \leq 1.01$).

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 (Figure 7; Table 5). 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.

Figure 7

Perceptions of the Person and the Love Letter in Study 5



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. Coloured 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 5

Pairwise Contrasts in Study 5

	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.

Exploratory text analysis of participants' free responses supported this quantitative pattern (see Supplementary Materials for methodology and Supplementary Table 7 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 (Figure 8). 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 related to character evaluations, suggesting that effort works solely through perceptions of authenticity and care.

Discussion

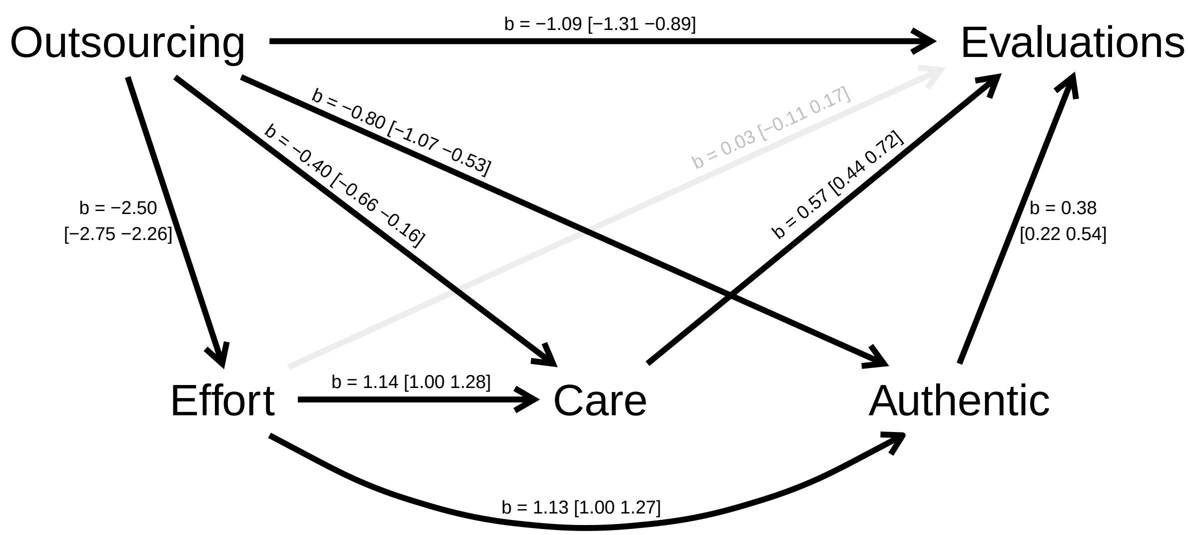
In our final study, we replicated our previous findings that outsourcing to AI results in negative character evaluations. Expanding on our previous studies, we also found that people spontaneously rate outsourcing to AI as less effortful, less authentic, and indicative of less care in the task. Our path model suggested that outsourcing to AI influences character evaluations in part through these mechanisms. According to our model, outsourcing to AI is seen as less effortful, which in turn signals a lack of authenticity and care in the task, resulting in negative character evaluations. The fact that perceived effort did not have a direct effect on character evaluations indicates that effort is important only insofar as it signals a lack of authenticity and care.

General Discussion

Across five pre-registered experiments, we have demonstrated negative perceptions of outsourcing to AI. People who outsourced tasks to AI were perceived more negatively than people who completed the tasks by themselves. These negative impressions were particularly strong for

Figure 8

Path Model in Study 5



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).

people who used AI to complete socio-relational tasks, such as writing a love letter or writing wedding vows, and for people who copied the model's first output verbatim without acknowledging their reliance on AI. Outsourced work was also perceived as less meaningful, less authentic, and less reward-worthy than ostensibly human-generated writing. Our experiments suggested that these negative evaluations are driven in part by a perceived lack of effort, a perceived lack of care in the task, and a perceived reduction in authenticity of the outsourced work.

The findings in this paper build on prior research in a number of ways. Our results corroborate and advance research on the moralisation of effort. Studies have shown that people inherently value effort (Inzlicht et al., 2018; Kruger et al., 2004) and perceive displays of effort as costly signals of one's moral character and cooperative intent (Celniker et al., 2023; Cubitt et al., 2011; Tissot & Roth, 2025). However, prior research has not explored the mechanisms by which perceptions of effort influence character evaluations – what exactly is effort signalling? We have suggested that the reduced effort from outsourcing socio-relational tasks to AI signals that the work is less authentically their own and that the person cares less about the task (and therefore, perhaps, the relationship). The lack of a direct effect of perceived effort in our path model showed that it is inferences of authenticity and care, 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 results also contribute to the growing literature on algorithm aversion. Research has shown that people tend to distrust AI models, even when they show identical or superior performance to humans (Dietvorst et al., 2015). This aversion is particularly strong in the moral domain (Bigman & Gray, 2018). Building on this work, we have shown that the aversion to algorithms can spill over into negative evaluations of people who unthinkingly copy the outputs of algorithms, especially for socio-relational tasks.

Regarding deception in AI use, our findings are relevant for the discussion about acknowledging the use of AI in jobs that require writing, such as journalism (Cools & Diakopoulos, 2024) and academia (Marescotti, 2023). Many scientific journals, for example, now require that authors declare their use of generative AI. However, thus far, there appears to be an AI attribution gap: many authors anonymously admit using AI to write scientific papers, but only few actually declare it in their published work (Gignac, 2025). Our findings make the case for honestly declaring use of AI. Such honesty could increase trust in the authors, and perhaps in the work itself.

Beyond the studies presented here, our results raise further questions about outsourcing to AI that should be addressed in future research. First, the fact that we found a direct effect of outsourcing on character evaluations in our final study, even after accounting for the indirect effects of effort, authenticity, and care, suggests that there are alternative mechanisms underlying the negative perceptions of outsourcing to AI that remain unexplored. Such mechanisms might plausibly include the perceived lack of creativity and originality of AI-generated work, the generic nature of the writing and lack of specificity to the intended recipient, questions around personal authorship and sincerity, and a potential lack of critical thinking when using AI (Lee et al., 2025). Future research should test these alternative mechanisms and build on the path model in Figure 8.

Second, we did not focus on individual differences in this paper, but it is reasonable to expect that people might vary in their reactions to AI-outsourcing. For example, research has shown that people high in openness to experience are more accepting of emerging technologies (Devaraj et al., 2008; Watjatrakul, 2016). People high in openness might therefore perceive outsourcing less negatively. Conversely, people with higher AI literacy and who are more resistant to AI on moral grounds tend to have more negative attitudes towards AI (de Mello et al., 2025; Tully et al., 2025) and so may perceive AI-outsourcing more negatively. Future work should explore whether these individual differences moderate perceptions of outsourcers.

Third, while we have demonstrated negative perceptions of outsourcing in this paper, it remains unclear if and 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, research in biopsychology has found that people are more accepting of cognitively-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.

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 generalisability of our findings, we focused specifically on participants from the United Kingdom, a relative global outlier in its high levels of AI readiness and development (Oxford Insights, 2024; Tortoise Media, 2024). We also collected data at a time when AI is still an emerging technology. It remains unclear whether our results will generalise to other countries with varying levels of AI infrastructure and whether negative perceptions of outsourcers will persist as AI use becomes more commonplace in the coming years. Another limitation is our reliance on hypothetical vignettes to assess people's perceptions of outsourcing to AI. It could be argued that these vignettes lack the richness of information to make informed character evaluations and do not realistically capture the way that AI is actually being used "in the wild". Our inclusion of a tool outsourcing condition somewhat mitigates these concerns, but it would be interesting to explore how people react to real-world examples of outsourcing.

In sum, we have demonstrated negative perceptions of outsourcing to AI. People perceive individuals who outsource tasks to AI more negatively across a range of character dimensions and perceive outsourced work as less meaningful and authentic. These negative perceptions are particularly strong for socio-relational tasks, such as writing wedding vows, and are compounded when the outsourcer copies the AI's output verbatim and does not honestly acknowledge their use of AI. Taken together, our findings suggest that, in a world of algorithm-mediated interactions, AI is no substitute for investing effort into our interpersonal relationships.

References

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., Dervieux, C., & Woodhull, G. (2024). *Quarto* (Version 1.7.23) [Computer software]. https://doi.org/10.5281/zenodo.5960048

Barcaui, A., & Monat, A. (2023). Who is better in project planning? Generative artificial intelligence or project managers? *Project Leadership and Society*, *4*, 100101. https://doi.org/10.1016/j.plas.2023.100101

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

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

Cools, H., & Diakopoulos, N. (2024). Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities. *Journalism Practice*, 1–19. https://doi.org/10.1080/17512786.2024.2394558

Cubitt, R. P., Drouvelis, M., Gächter, S., & Kabalin, 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

de Mello, V. O., Ayad, R., Côté, É., Inbar, Y., Plaks, J., & Inzlicht, M. (2025). *The moralization of artificial intelligence*. PsyArXiv. https://doi.org/10.31234/osf.io/5mwre_v2

Devaraj, S., Easley, R. F., & Crant, J. M. (2008). Research note: How does personality matter? Relating the five-factor model to technology acceptance and use. *Information Systems Research*, *19*(1), 93–105. http://www.jstor.org/stable/23015423

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

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

Gignac, G. E. (2025). The AI attribution gap: Encouraging transparent acknowledgment in the age of AI. *Intelligence*, *108*, 101880. https://doi.org/10.1016/j.intell.2024.101880

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

Kirk, C. P., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. *Journal of Business Research*, *186*, 114984. https://doi.org/10.1016/j.jbusres.2024.114984

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

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

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *CHI 2025*. https://www.microsoft.com/en-us/research/publication/ the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence

Marescotti, M. (2023). To ChatGPT or not to ChatGPT: The use of artificial intelligence in writing scientific papers. In *Brain Communications* (6; Vol. 5, p. fcad266). Oxford University Press US. https://doi.org/10.1093/braincomms/fcad266

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?* [Online forum post]. Reddit. https://www.reddit.com/r/relationship_advice/comments/1d9tw9j/my_30f_husband_36m_used_ai_to_write_his_vows_how/

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-imaginist.com

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The impact of AI on developer productivity: Evidence from GitHub copilot*. arXiv. https://arxiv.org/abs/2302.06590

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rainie, L., Funk, C., Anderson, M., & Tyson, A. (2022). *How Americans think about artificial intelligence*. Pew Research Center. https://www.pewresearch.org/internet/2022/03/17/how-americans-think-about-artificial-intelligence

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

Sedikides, C., & Schlegel, R. J. (2024). Distilling the concept of authenticity. *Nature Reviews Psychology*, *3*(8), 509–523. https://doi.org/10.1038/s44159-024-00323-y

Silver, I., Newman, G., & Small, D. A. (2021). Inauthenticity aversion: Moral reactance toward tainted actors, actions, and objects. *Consumer Psychology Review*, *4*(1), 70–82. https://doi.org/10.1002/arcp.1064

Tissot, T., & Roth, L. H. O. (2025). *Is it worth the hustle? A multi-country replication of the effort moralization effect and an extension to generational differences in the appreciation of effort*.

OSF Preprints. https://doi.org/10.31219/osf.io/ck4st_v5

Tortoise Media. (2024). *The global AI index*. https://www.tortoisemedia.com/data/global-ai

Tully, S. M., Longoni, C., & Appel, G. (2025). Lower artificial intelligence literacy predicts greater AI receptivity. *Journal of Marketing*. https://doi.org/10.1177/00222429251314491

Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*, *42*(1), 69–77. https://doi.org/10.1016/j.jesp.2005.02.003

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

Watjatrakul, B. (2016). Online learning adoption: Effects of neuroticism, openness to experience, and perceived values. *Interactive Technology and Smart Education*, *13*(3), 229–243. https://doi.org/10.1108/ITSE-06-2016-0017

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Supplementary Materials

Negative Perceptions of Outsourcing to Artificial Intelligence

Scott Claessens, Pierce Veitch, and Jim Everett

School of Psychology, University of Kent

Pilot Study 1

Methods

Participants

We recruited a convenience sample of 200 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 186 participants (118 female; 67 male; 1 non-binary / third gender; 0 undisclosed gender; mean age = 38.99 years).

Procedure

We presented participants with six different tasks "that people might perform in their daily lives". 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 task, we asked participants the following questions on 7-point Likert scales:

- Is this a social task?
- Does this task require social skills?
- Does this task impact other people?
- How important are the consequences of this task?
- How important is it that effort goes into this task?
- How important is it that others see the effort that goes into this task?

Statistical Analysis

We fitted a Bayesian multivariate multilevel cumulative-link ordinal model to the data using the *brms* R package. We modelled each task evaluation as a separate response variable and included correlated varying intercepts for participants and tasks. We used regularising priors for all parameters to impose conservatism on parameter estimates (see Supplementary Materials for full model specification). The model converged normally ($\hat{R} \leq 1.01$).

Results

We found that participants' responses to all six questions tended to be positively correlated. For example, tasks rated as more social were also rated as requiring more social skills (see Supplementary Figure 15). Estimated averages and rankings for the 20 tasks across each of the questions can be found in Supplementary Figures 16 – 21.

Pilot Study 2

Methods

Participants

We conducted a power simulation to determine our target sample size. 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 Prolific. 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 (see Supplementary Figure 22).

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.

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 (see Supplementary Figure 22).

Pre-registration

We pre-registered the study on the Open Science Framework (https://osf.io/khr42).

Statistical Analysis

We fitted Bayesian multivariate multilevel cumulative-link ordinal models to the data using the *brms* R package. We modelled each character evaluation – competence, warmth, morality, laziness, and trustworthiness – as a separate response variable and included fixed effects for conditions, varying intercepts for participants, and varying intercepts and slopes for tasks. We used regularising priors for all parameters to impose conservatism on parameter estimates (see Supplementary Materials for full model specifications). All models converged normally ($\hat{R} \leq 1.01$).

Results

We found that people who outsourced tasks to AI or other humans were perceived more negatively than people who completed the tasks themselves (Supplementary Figure 23). In particular, people who outsourced were perceived as lazier and less competent, with smaller yet detectable differences for perceptions of warmth, morality, and trustworthiness (Supplementary Table 8). 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 (Supplementary Figure 24). 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 the first pilot study. 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 Figures 25 and 26). In other words, outsourcing to AI or other humans is perceived more negatively for tasks that have these features, compared to tasks without these features.

Vignette Wording in Study 5

We presented participants in Study 5 with the following vignette text:

> Adam has been dating his partner for almost a year, and Valentine's Day is coming up. He knows that many people exchange a card on Valentine's Day containing a love letter to their partner, and he decides to send a love letter to his partner too.
>
> This year, Adam has been closely following developments in technology and has read of people using AI tools like ChatGPT for things like this, either using it to help with writing or getting AI to do the task completely.

This was followed by the manipulation text (see main text). Participants were then presented with the love letter that Adam ostensibly wrote, which was held constant across conditions:

> Happy Valentine's Day, my love.
>
> I don't think I tell you enough just how much you mean to me. Being with you feels like breathing a little easier, like the world is a bit softer just because you're in it. You make the everyday feel special, and somehow you always know how to calm my nerves or make me laugh at just the right moment. I feel like myself with you – maybe even a better version of myself – and that's such a rare and beautiful thing.
>
> I'm so grateful for you – for the way you listen, the way you love, the way you show up, even in the small ways. I hope you know that no matter what, I'm always in your corner. I can't wait to keep making memories together, whether we're off on some adventure or just curled up on the couch. I love you more than I can really put into words, but I promise I'll spend every day trying.
>
> Yours,
>
> Adam

Methods for Text Analysis in Study 5

To generate frequency lists for each experimental condition in Study 5, we created three documents containing the raw text submissions to the open-ended question “In your own words, describe how you feel about Adam and why”. Each raw text submission was paired with a numbered text ID column. The number of submissions was roughly equivalent across conditions: the control condition (N = 196), the tool outsourcing condition (N = 215), and the full outsourcing condition (N = 202).

All text processing was conducted using the Basic Unit-Transposable Text Experimentation Resource (BUTTER; Version 0.9.4.1; Boyd, 2019). To prepare the data, each CSV file was converted into a folder containing individual text files – one per submission – using two plugins: *Read Text from CSV* (Version 1.0.2) and *Save .txt Files to Folder* (Version 1.0.6). The settings for *Read Text from CSV* were as follows: file encoding = UTF-8, row identifier = ID, text column = Text, CSV delimiter = , and CSV quote = “.

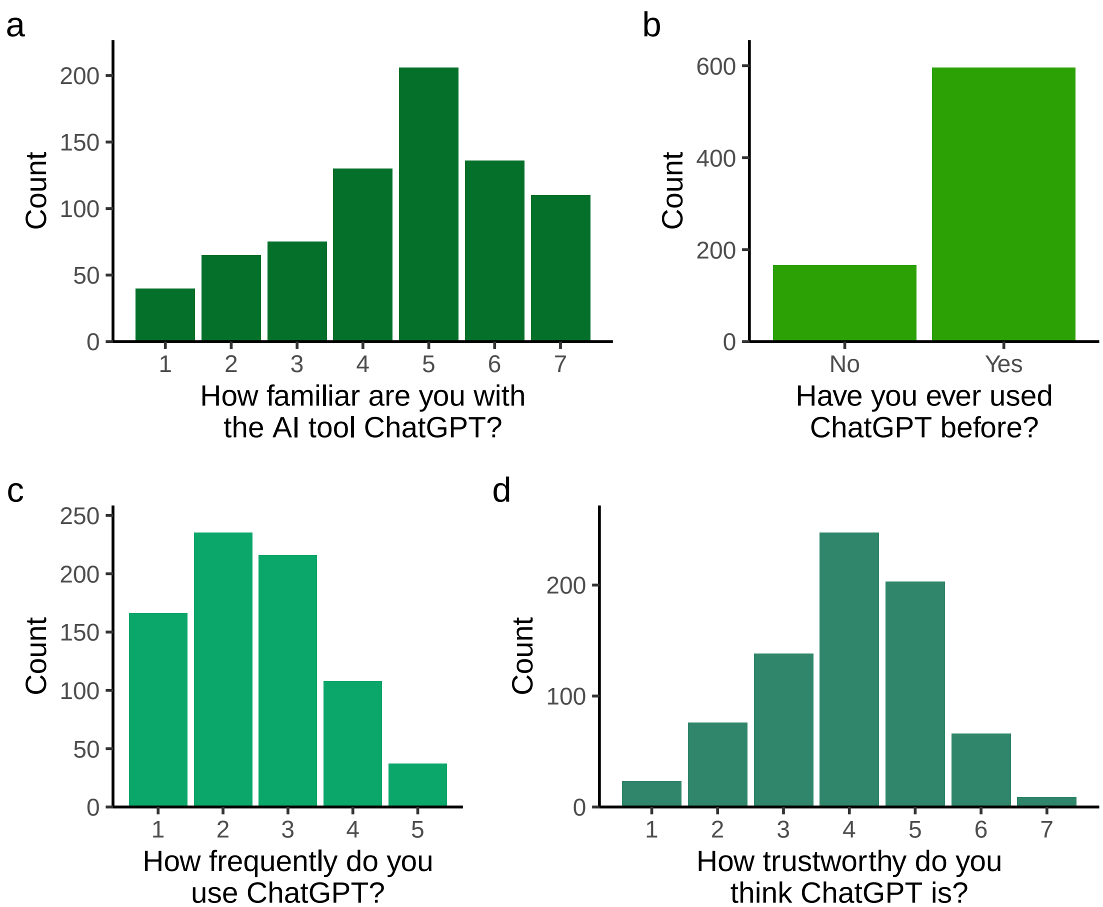
To generate frequency lists, we first loaded the .txt files using the *Load .txt Files from Folder* plugin (Version 1.0.4). Tokenization was performed using the Twitter-Aware Tokenizer (Version 1.0.2), with the options *convert text to lowercase* and *reduce elongation* enabled to minimize superficial variation in tokens. We removed filler and function words using the *Remove Stop Words* plugin (Version 1.0.31), applying the default English stop word list.

Frequency lists were created with the *Frequency List* plugin (Version 1.0.11). Settings included: unigram analysis (N = 1), omission of n-grams with frequency < 5, exclusion of n-grams appearing in fewer than 0.1% of documents, filtering collocates by Normalized Pointwise Mutual Information (NPMI), and removal of collocates with metric values < 0.5. Outputs were saved using the *Save Output to CSV* plugin (Version 1.0.5). This process was repeated separately for each condition folder.

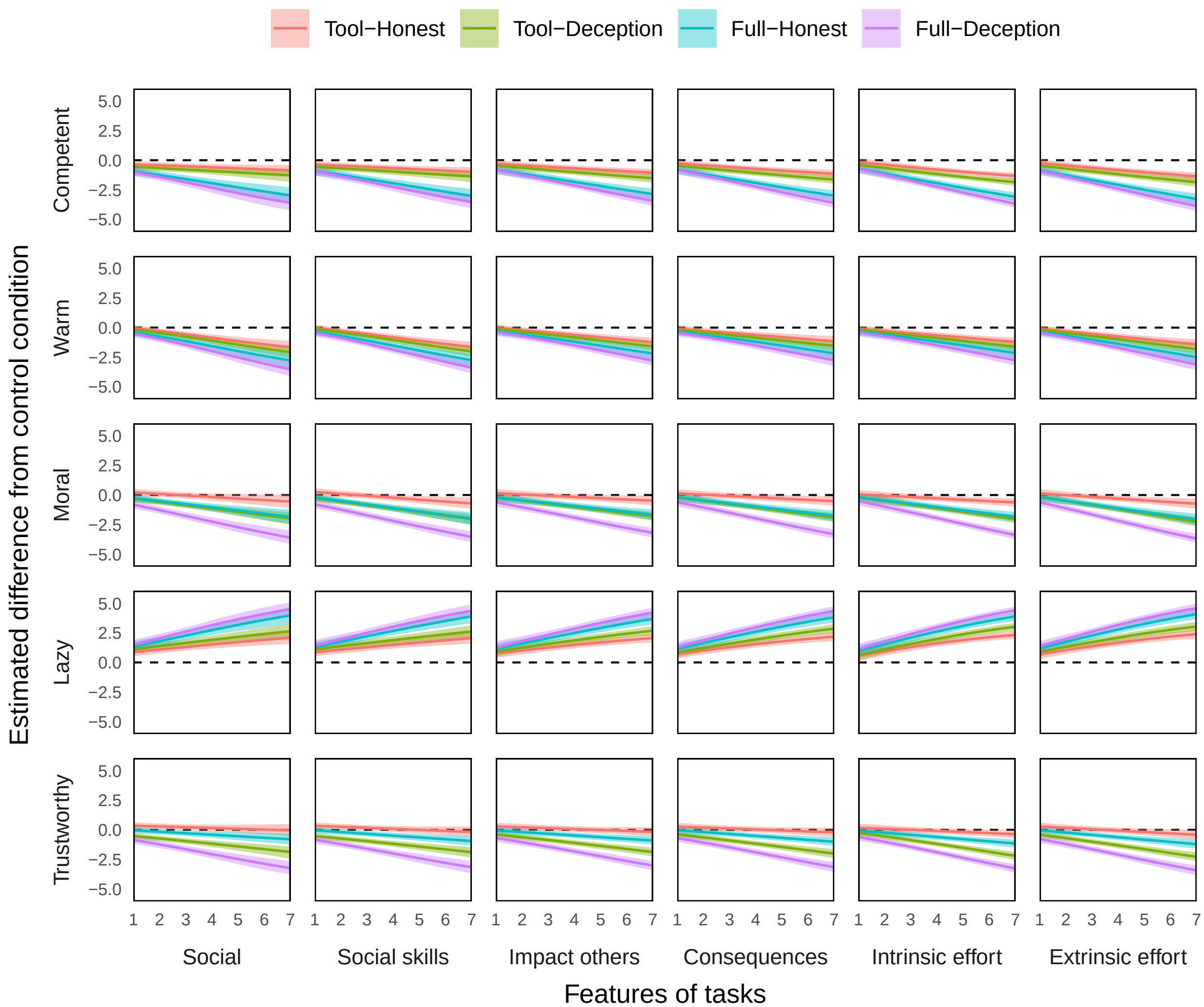
For cross-condition comparison, we used the *Compare Frequencies* plugin (Version 1.1.02), retaining most default settings. The only modification was disabling the *Skip comparisons with 0 frequency values* option. This plugin calculates a range of comparative metrics, including log likelihood (LL), %DIFF, Bayes Information Criterion (BIC), relative risk (RRisk), log ratio, and odds ratio.

Following previous work (e.g., Rayson & Garside, 2000; Gregson et al., 2022), we interpret %DIFF as an indicator of effect size and direction. Frequentist statistical significance was determined using log likelihood values, with the following thresholds: LL ≥ 3.84 ($p < .05$), LL ≥ 6.63 ($p < .01$), LL ≥ 10.83 ($p < .001$), and LL ≥ 15.13 ($p < .0001$).

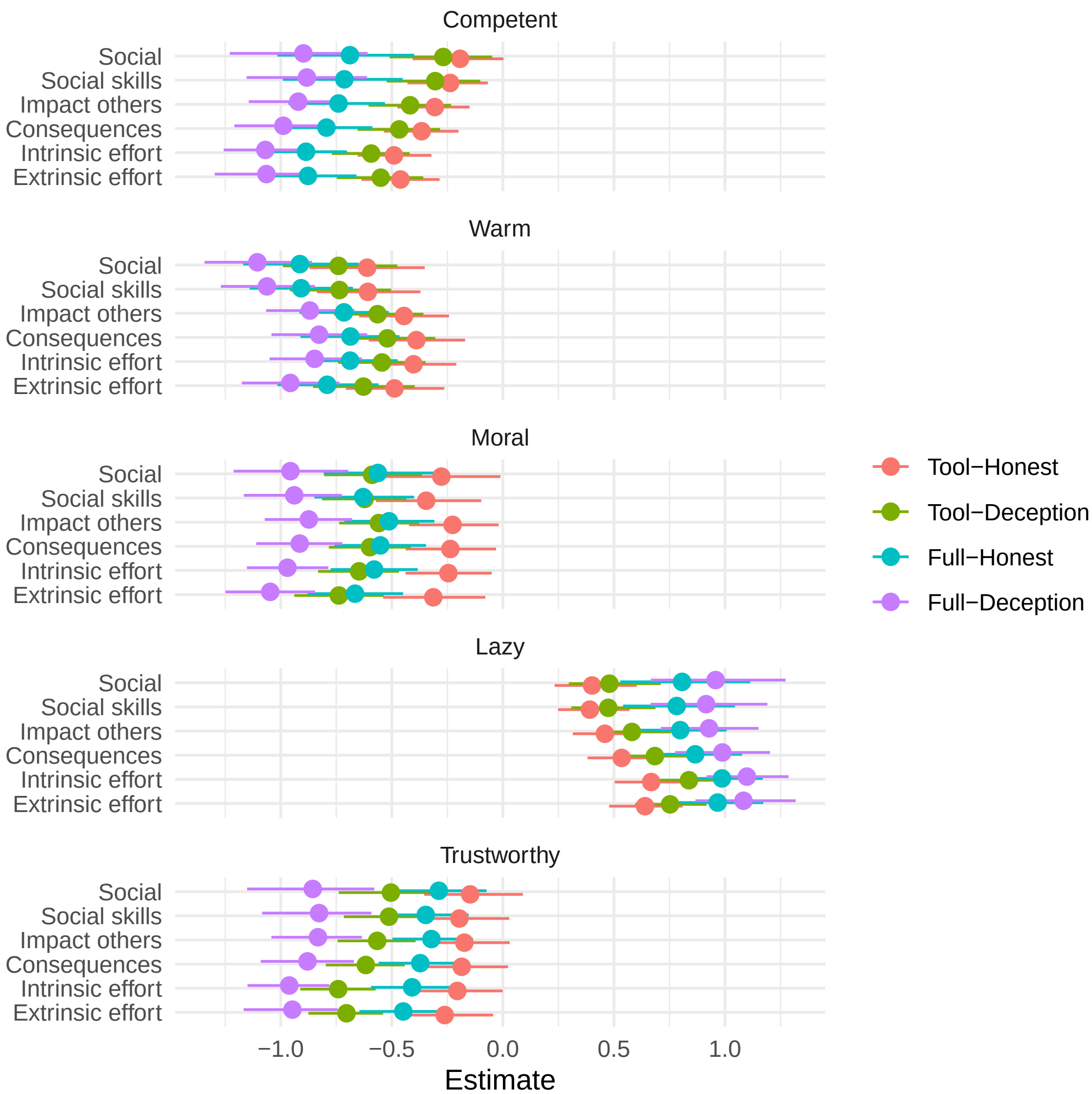
Supplementary Figures



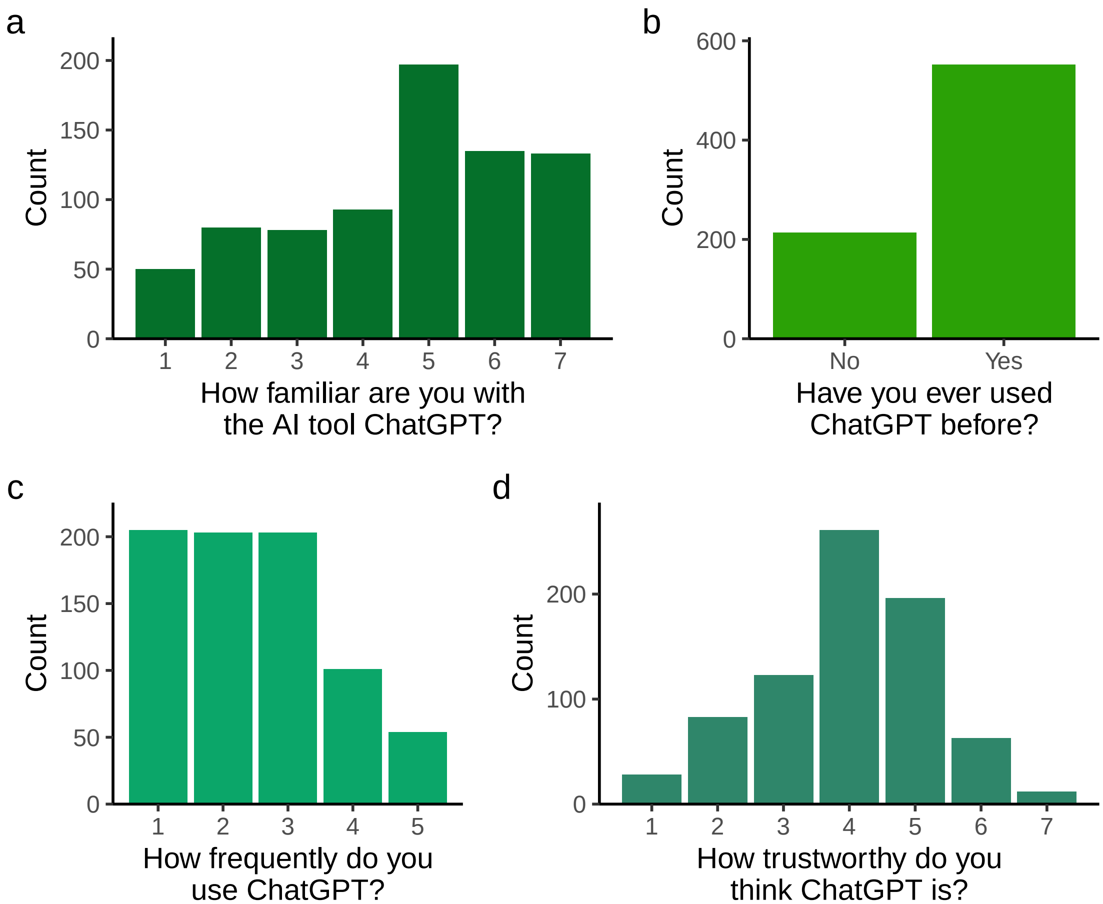
Supplementary Figure 1: Responses to the questions about ChatGPT in Study 1.



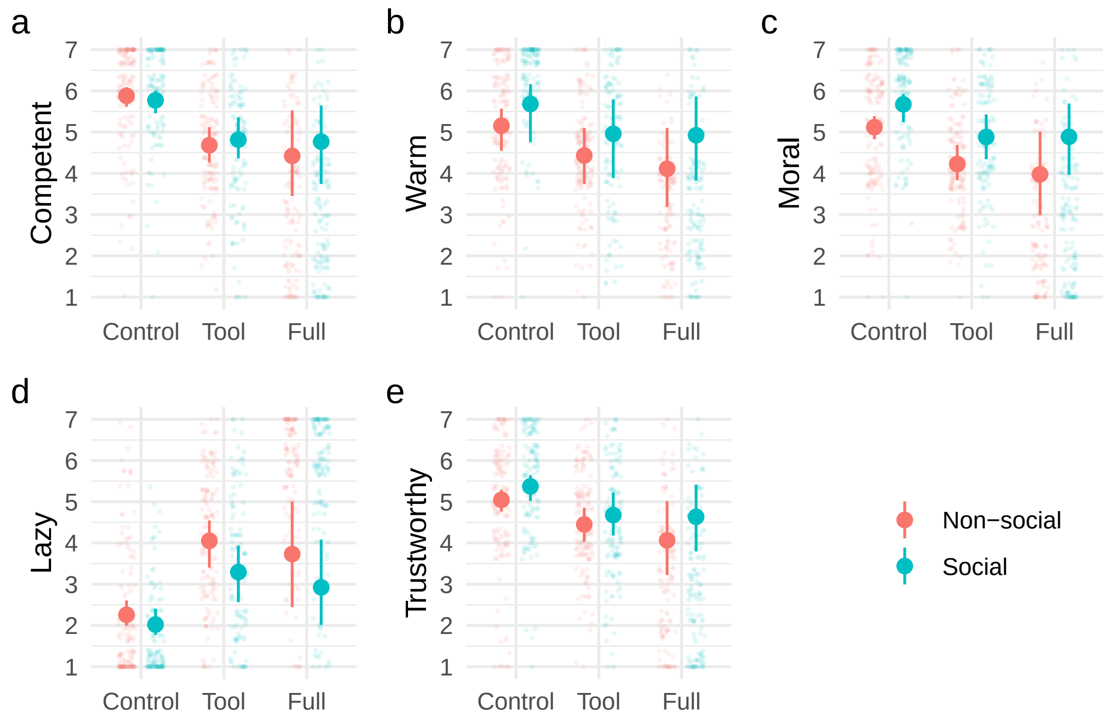
Supplementary Figure 2: The impact of task-specific features (e.g., being a social task) on the causal effects of outsourcing to AI compared to the control condition in Study 1. The y-axis reflects the estimated differences between the experimental conditions and the control condition (dashed line) on a 7-point Likert scale. Lines and shaded areas represent posterior medians and 95% credible intervals, respectively. The patterns indicate, for example, more negative effects of outsourcing on character evaluations for tasks that are rated as more social.



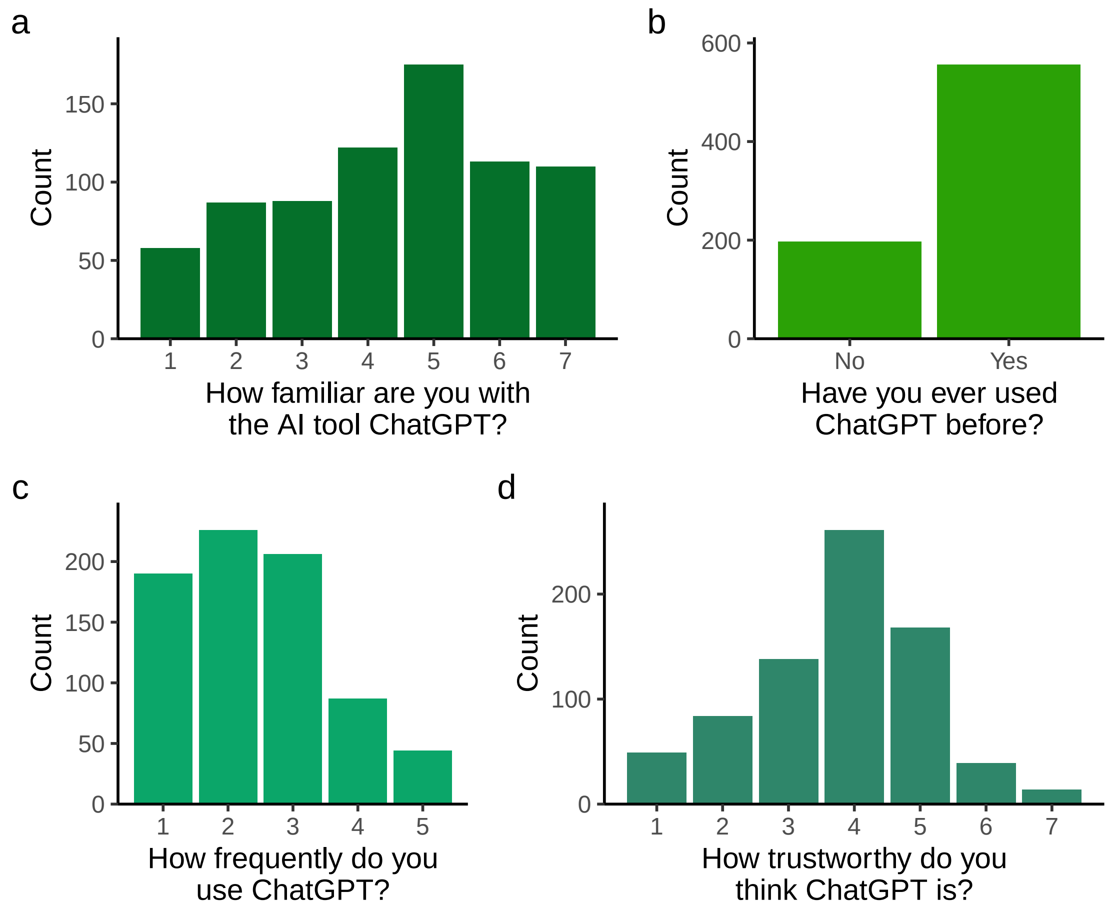
Supplementary Figure 3: Interaction parameters from models including task-specific features as moderators of the causal effects of AI outsourcing compared to the control condition in Study 1. Points and line ranges represent posterior medians and 95% credible intervals, respectively.



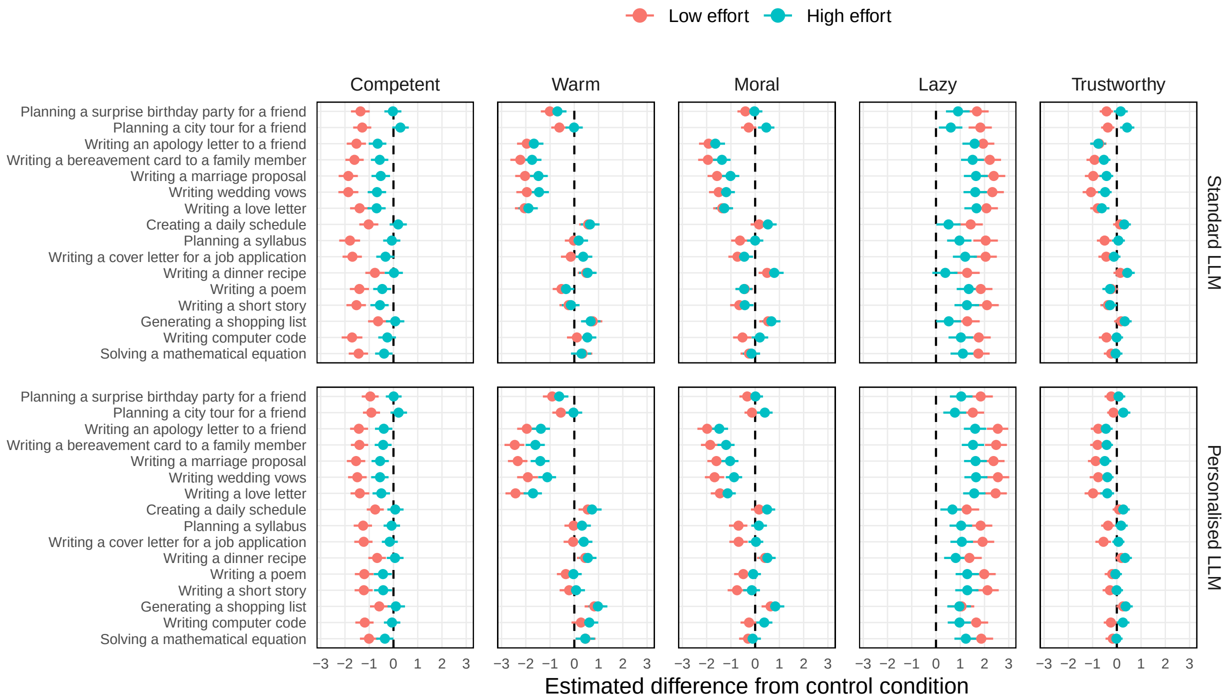
Supplementary Figure 4: Responses to the questions about ChatGPT in Study 2.



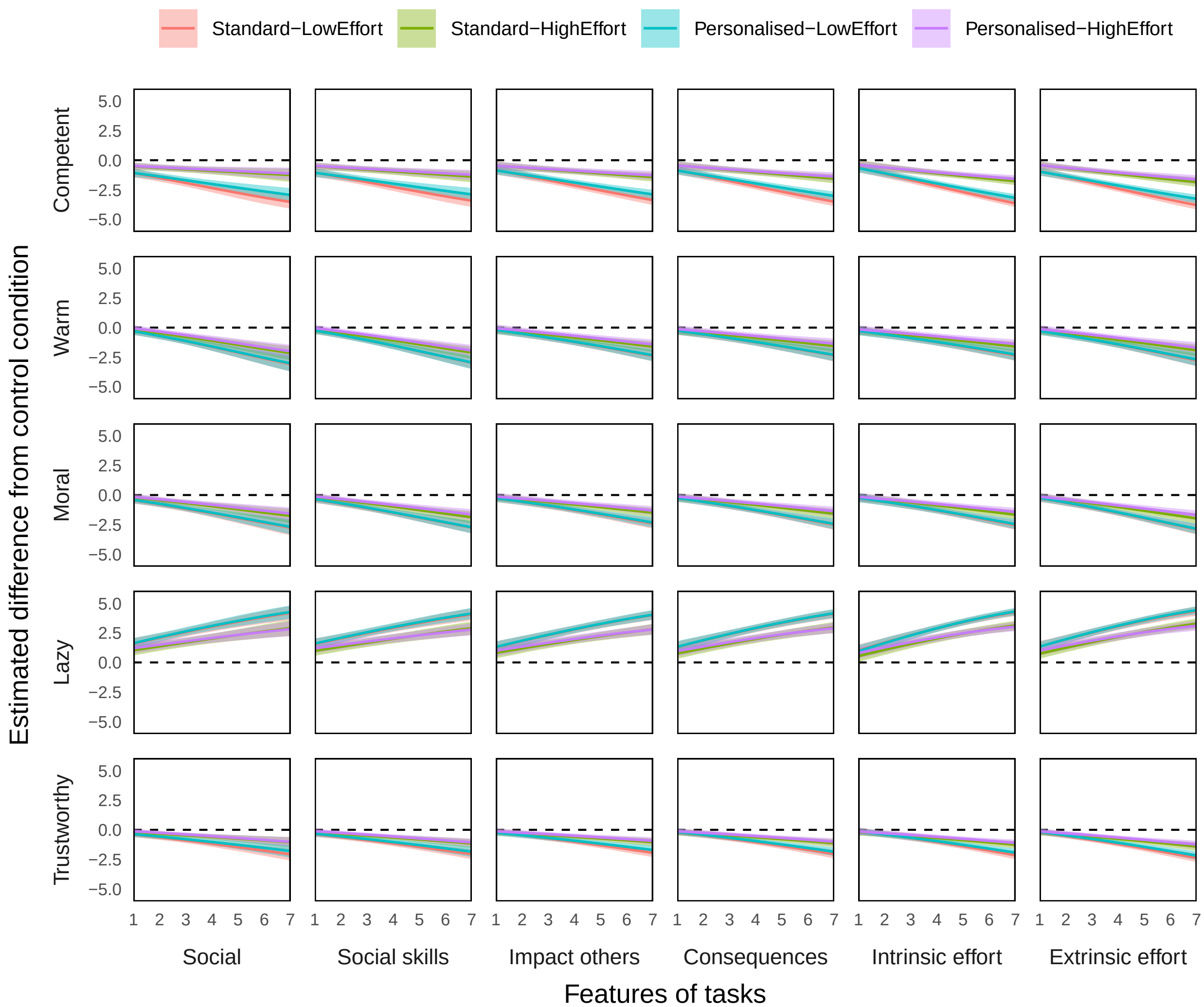
Supplementary Figure 5: Character evaluations in Study 2. Participants in the control condition, the tool outsourcing condition, and the full outsourcing condition evaluated the "other participant" on (a) competence, (b) warmth, (c) morality, (d) laziness, and (e) trustworthiness. Jittered points represent 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.



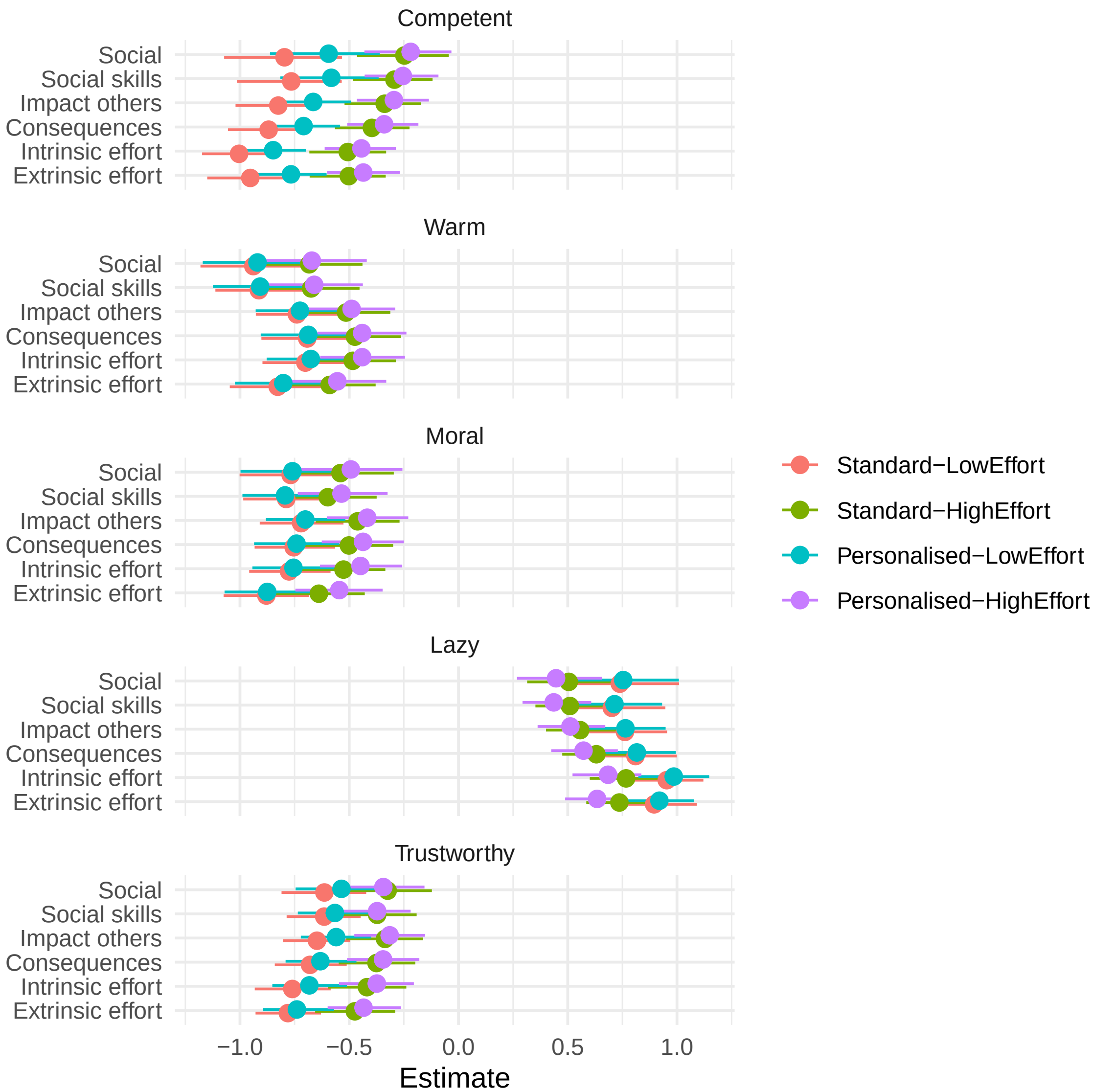
Supplementary Figure 6: Responses to the questions about ChatGPT in Study 3.



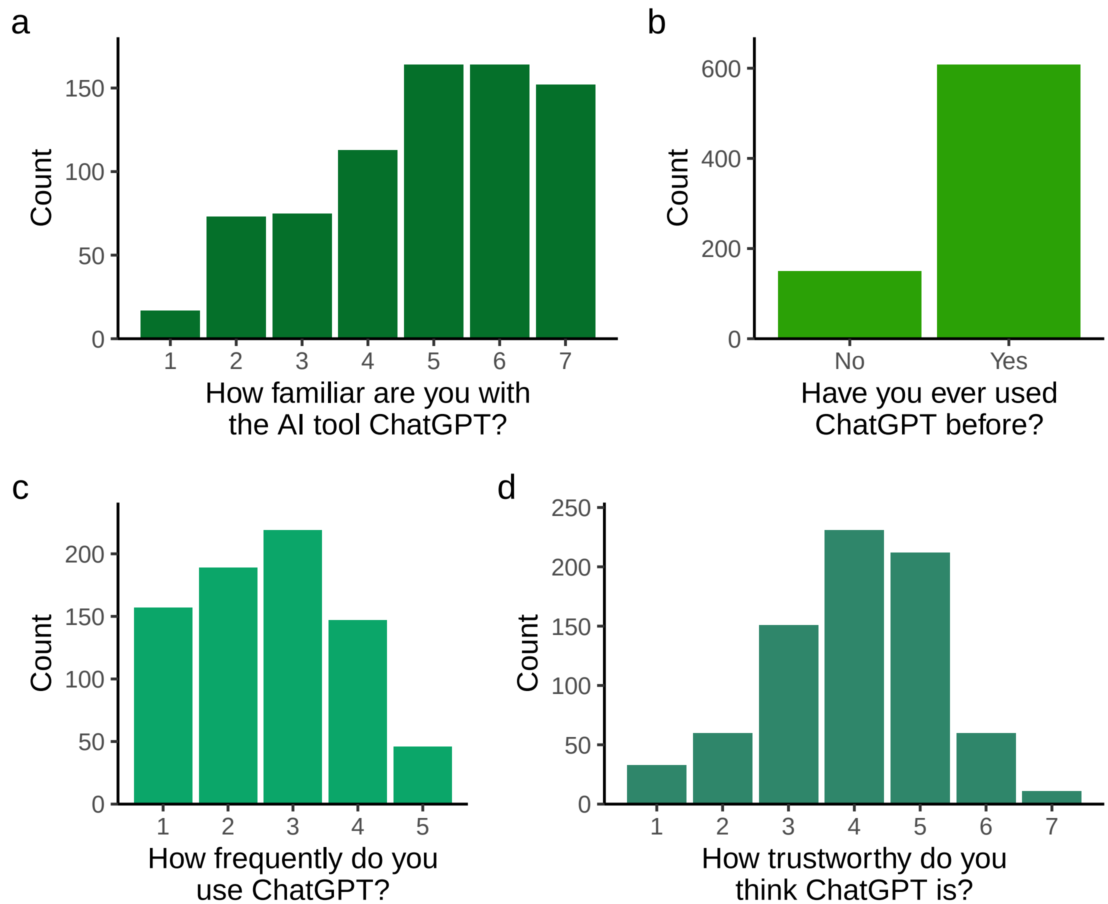
Supplementary Figure 7: Variation in the effects of outsourcing across tasks in Study 3. Tasks are ordered from most social (top) to least social (bottom) according to ratings from a pilot study. Point ranges are differences in marginal means on a 7-point Likert scale for the low effort conditions (red) and high effort conditions (blue) compared to the control condition. Upper panels refer to the standard LLM conditions, and lower panels refer to the personalised LLM conditions. Points and ranges represent posterior medians and 95% credible intervals, respectively.



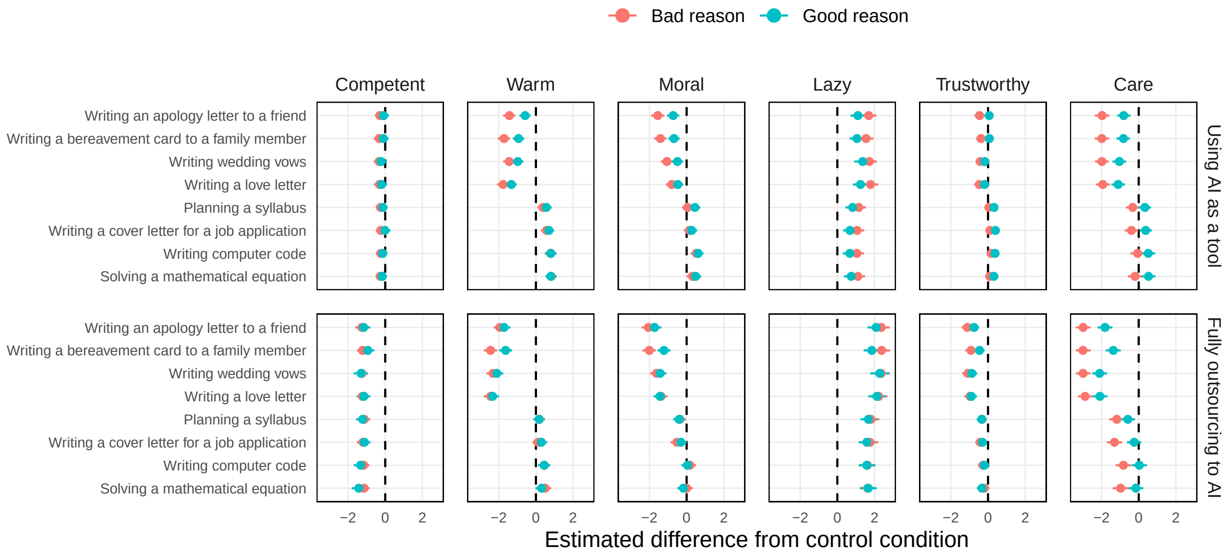
Supplementary Figure 8: The impact of task-specific features (e.g., being a social task) on the causal effects of outsourcing to AI compared to the control condition in Study 3. The y-axis reflects the estimated differences between the experimental conditions and the control condition (dashed line) on a 7-point Likert scale. Lines and shaded areas represent posterior medians and 95% credible intervals, respectively. The patterns indicate, for example, more negative effects of outsourcing on character evaluations for tasks that are rated as more social.



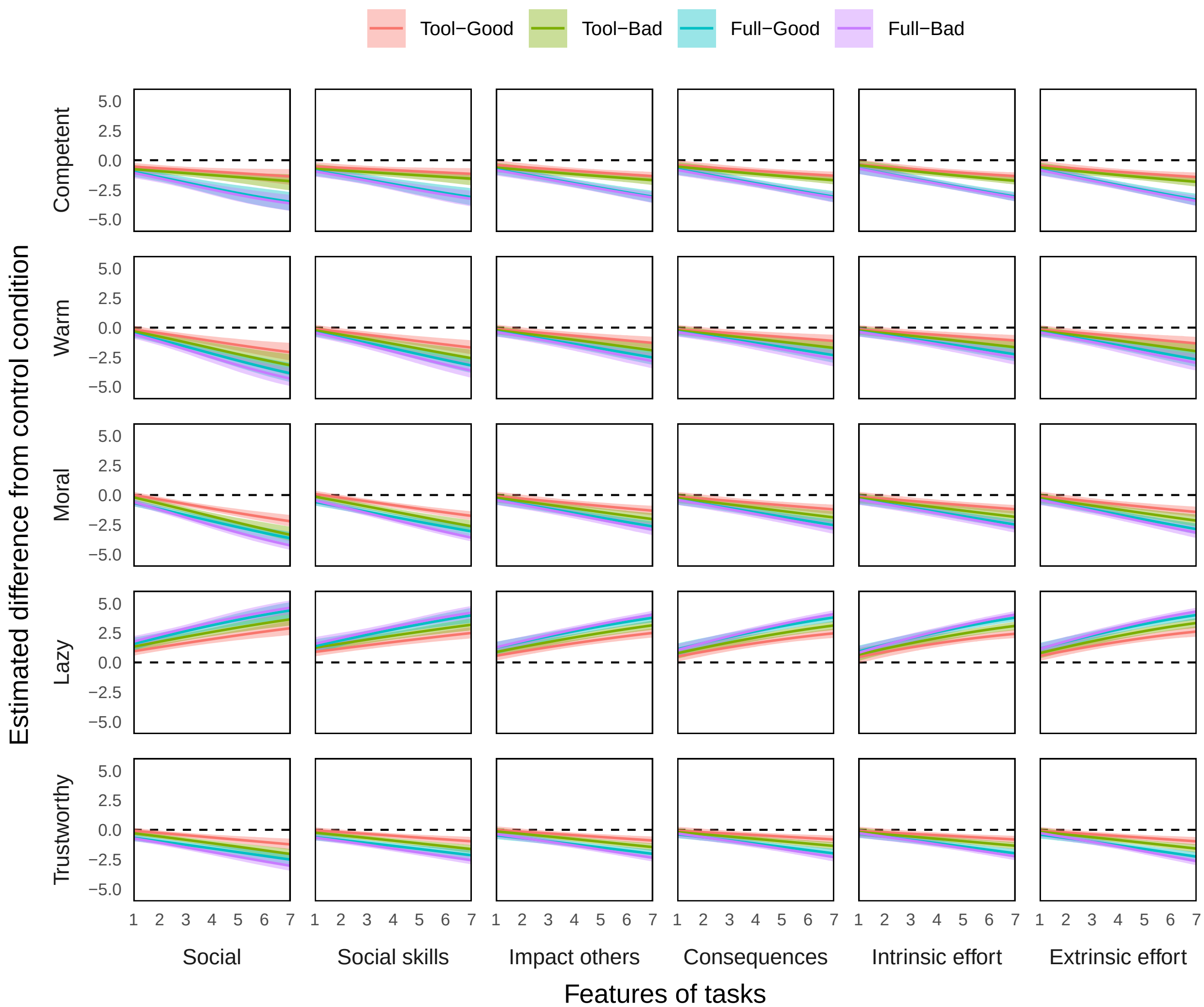
Supplementary Figure 9: Interaction parameters from models including task-specific features as moderators of the causal effects of AI outsourcing compared to the control condition in Study 3. Points and line ranges represent posterior medians and 95% credible intervals, respectively.



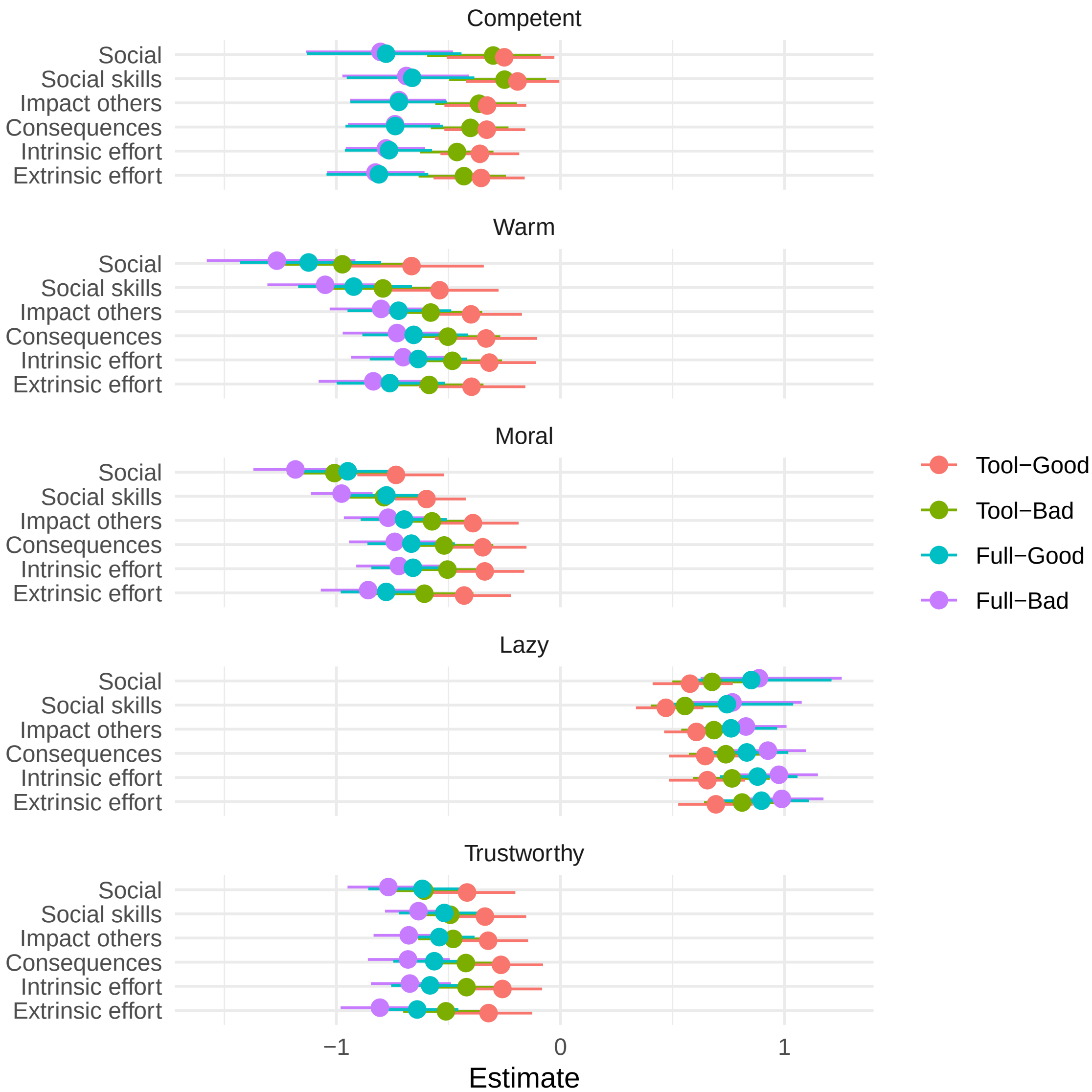
Supplementary Figure 10: Responses to the questions about ChatGPT in Study 4.



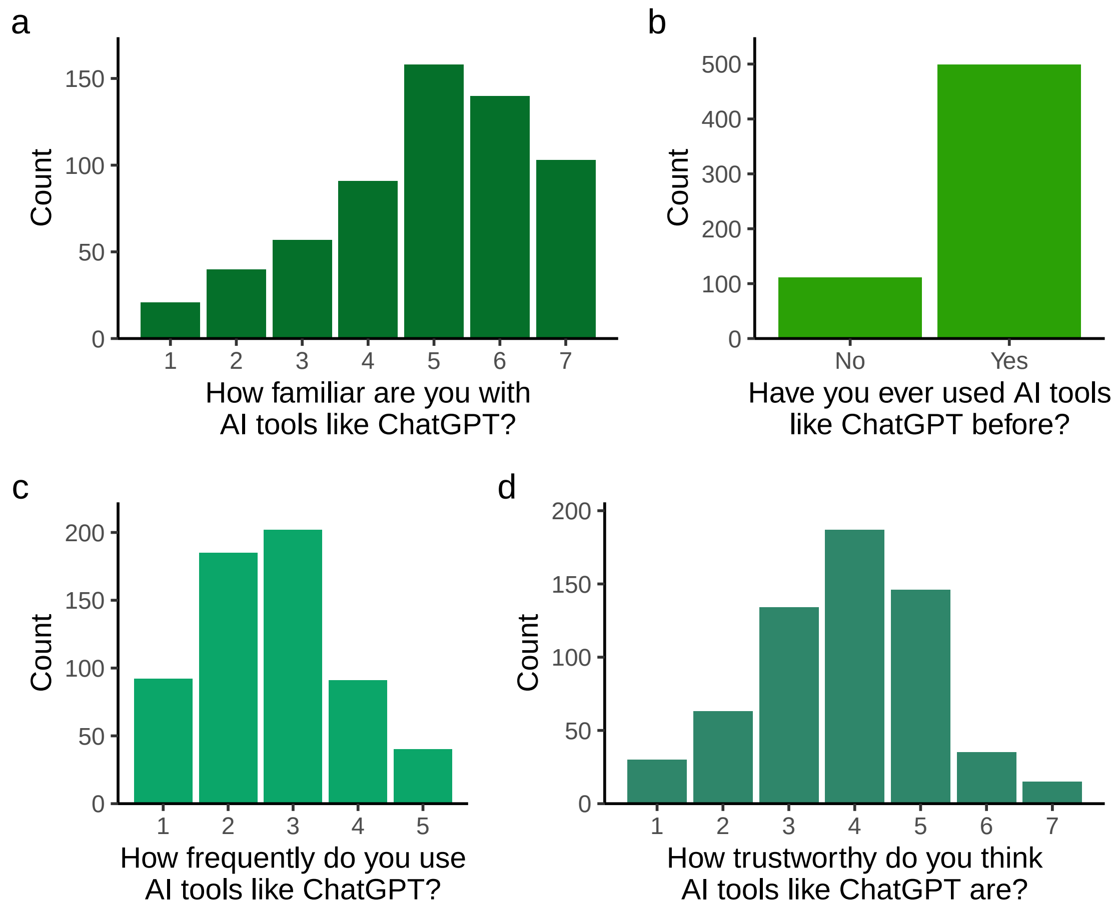
Supplementary Figure 11: Variation in the effects of outsourcing across tasks in Study 4. Tasks are ordered from most social (top) to least social (bottom) according to ratings from a pilot study. Point ranges are differences in marginal means on a 7-point Likert scale for the bad reason conditions (red) 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.



Supplementary Figure 12: The impact of task-specific features (e.g., being a social task) on the causal effects of outsourcing to AI compared to the control condition in Study 4. The y-axis reflects the estimated differences between the experimental conditions and the control condition (dashed line) on a 7-point Likert scale. Lines and shaded areas represent posterior medians and 95% credible intervals, respectively. The patterns indicate, for example, more negative effects of outsourcing on character evaluations for tasks that are rated as more social.

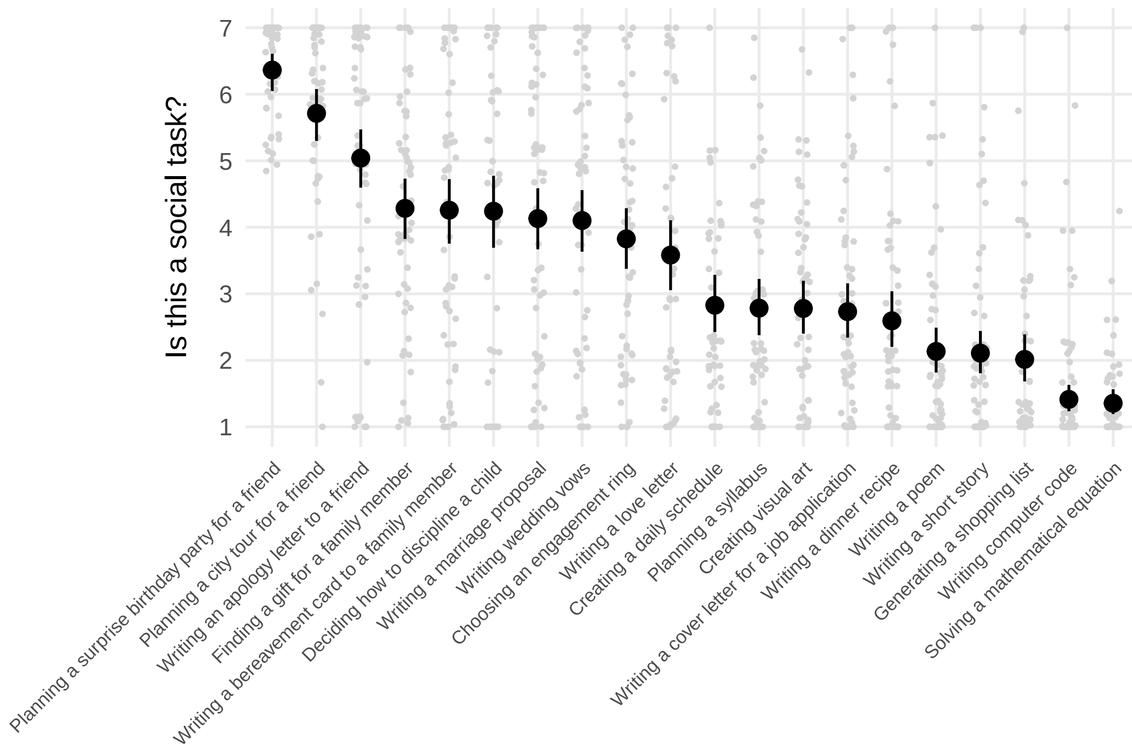


Supplementary Figure 13: Interaction parameters from models including task-specific features as moderators of the causal effects of AI outsourcing compared to the control condition in Study 4. Points and line ranges represent posterior medians and 95% credible intervals, respectively.

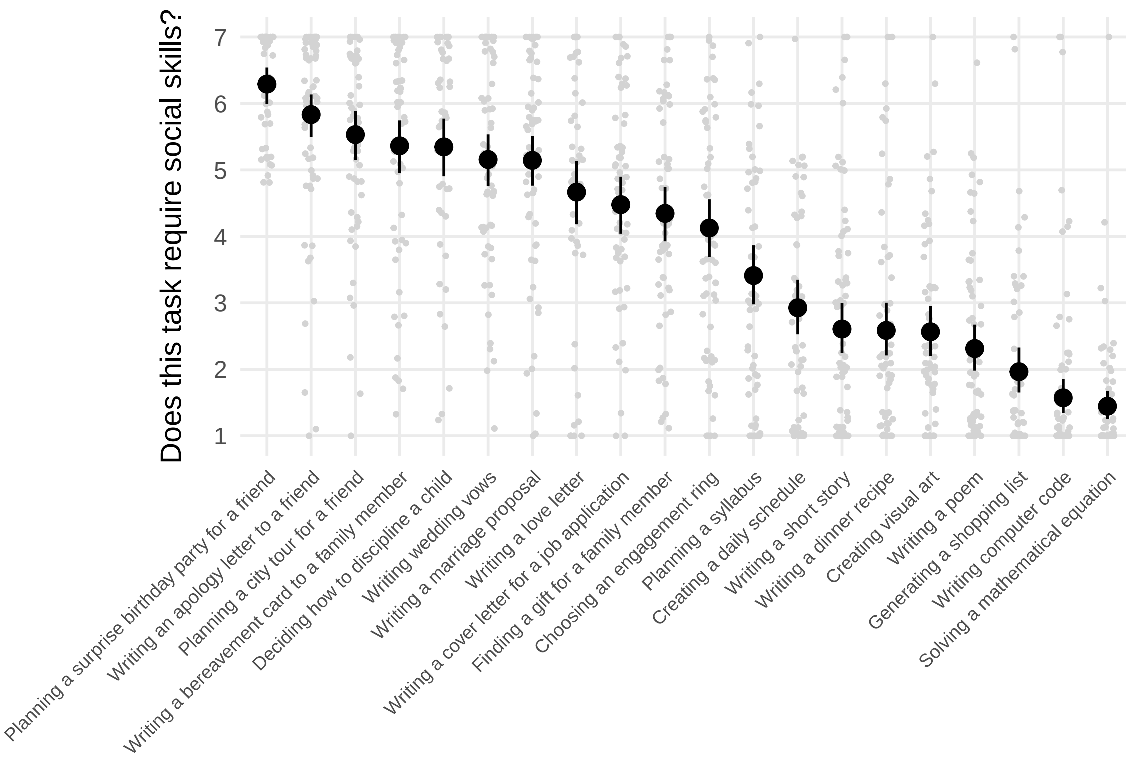


Supplementary Figure 14: Responses to the questions about ChatGPT in Study 5.

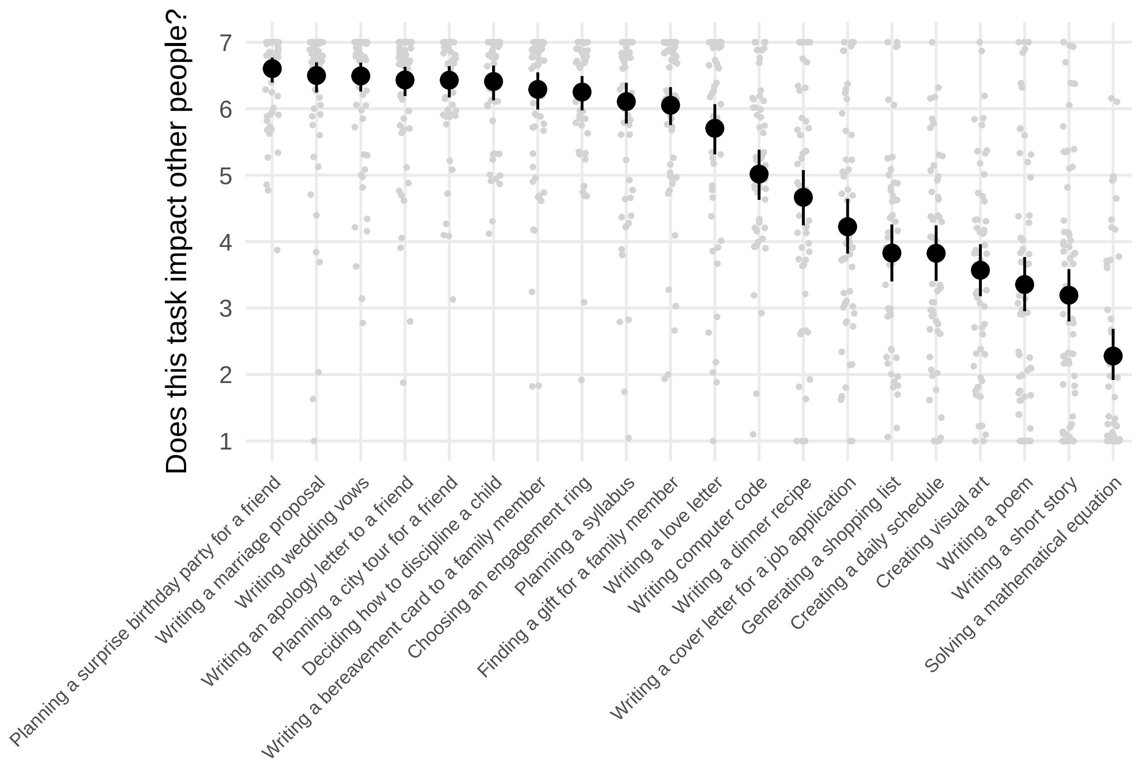
Supplementary Figure 15: Model-estimated task-specific correlations between all six questions in the first pilot study. Values are posterior median correlations. A positive correlation indicates that tasks that are rated highly on one question tend to be rated highly on another question.



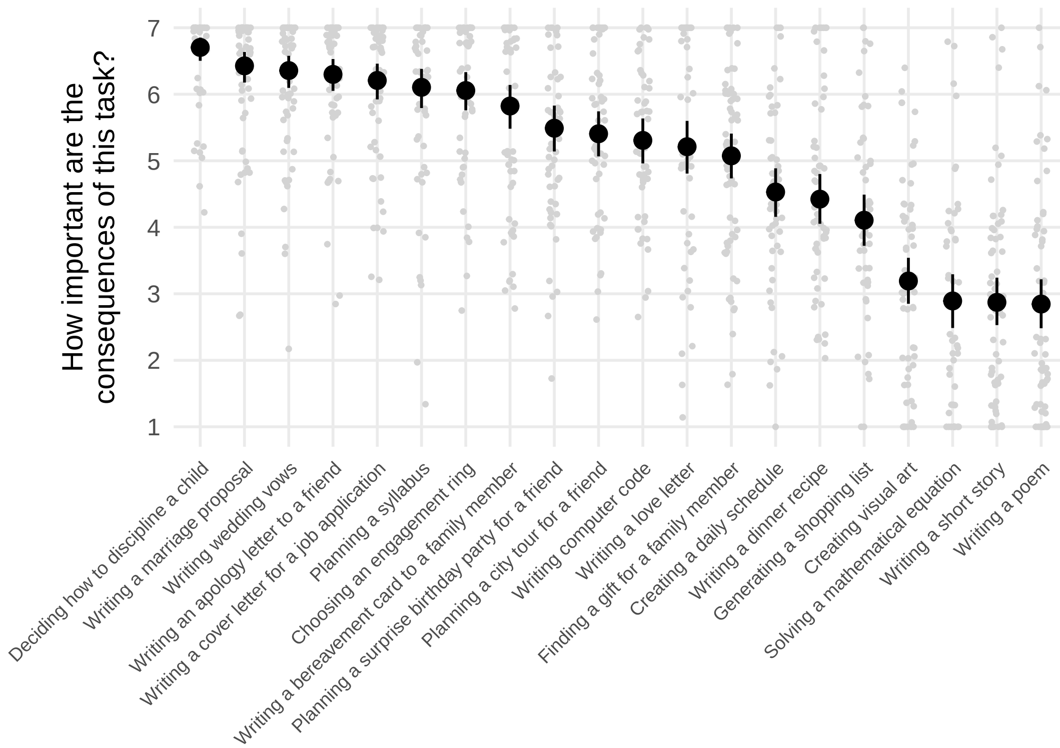
Supplementary Figure 16: Model-estimated means for the question “Is this a social task?” across all 20 tasks in the first pilot study. Grey points represent participant responses to the question, jittered for easier viewing. Black points are estimated means from the fitted model, pooling over participants. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.



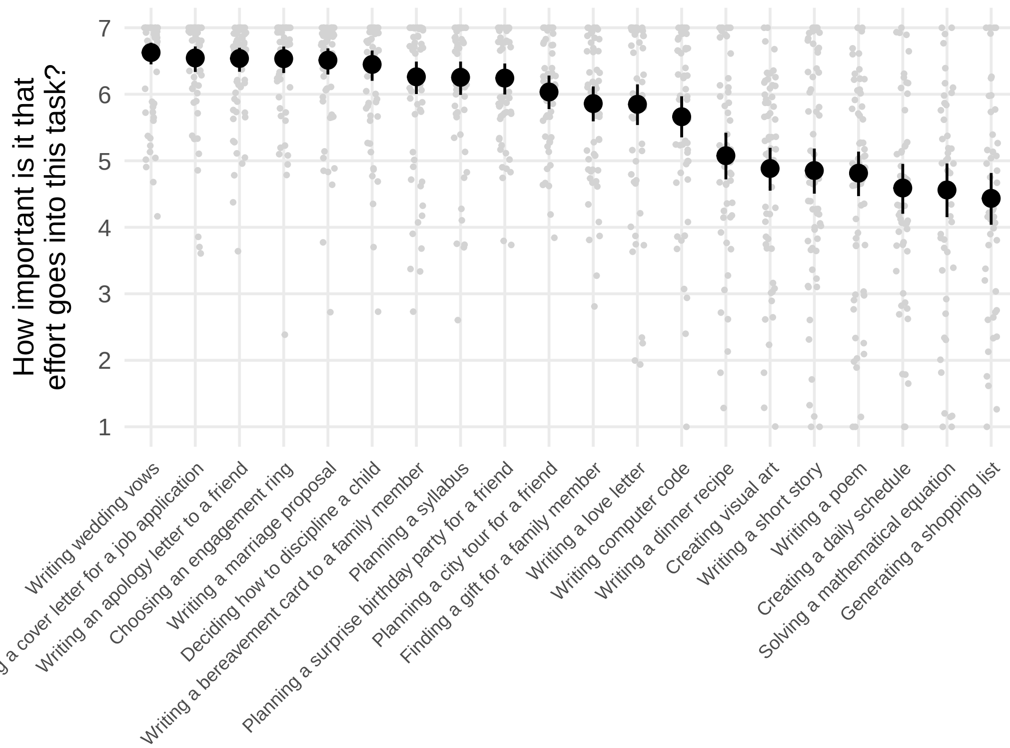
Supplementary Figure 17: Model-estimated means for the question "Does this task require social skills?" across all 20 tasks in the first pilot study. Grey points represent participant responses to the question, jittered for easier viewing. Black points are estimated means from the fitted model, pooling over participants. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.



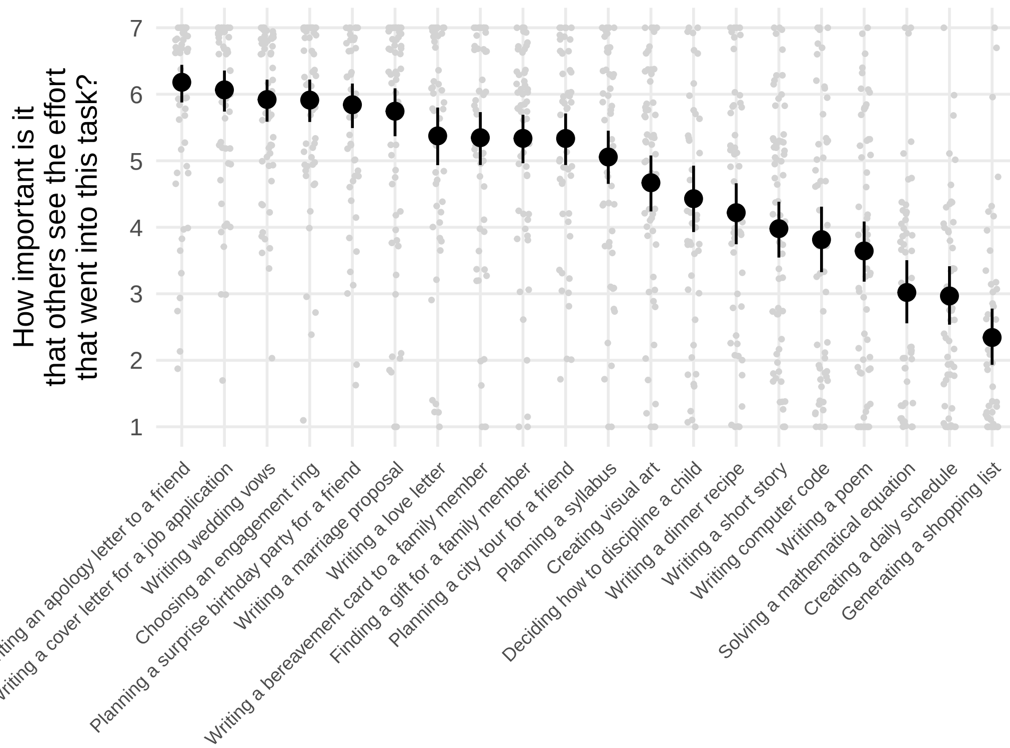
Supplementary Figure 18: Model-estimated means for the question "Does this task impact other people?" across all 20 tasks in the first pilot study. Grey points represent participant responses to the question, jittered for easier viewing. Black points are estimated means from the fitted model, pooling over participants. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.



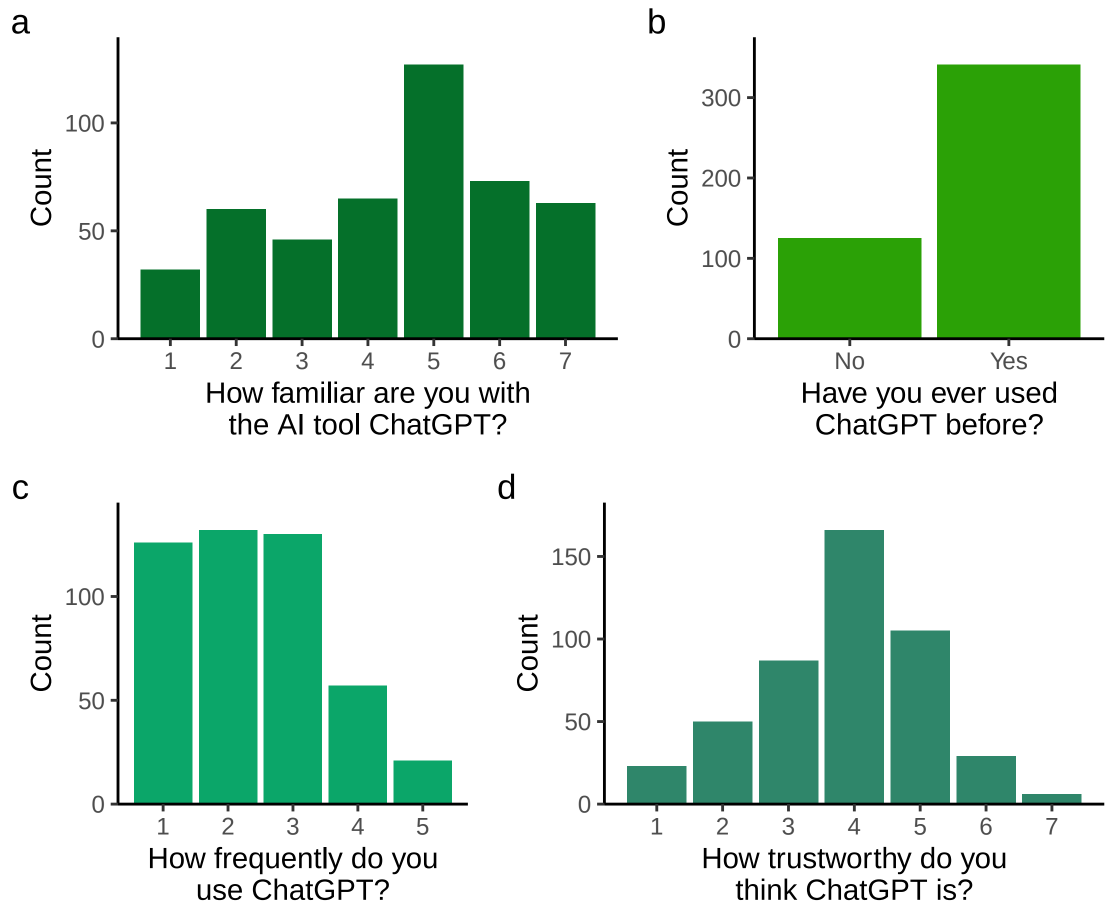
Supplementary Figure 19: Model-estimated means for the question "How important are the consequences of this task?" across all 20 tasks in the first pilot study. Grey points represent participant responses to the question, jittered for easier viewing. Black points are estimated means from the fitted model, pooling over participants. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.



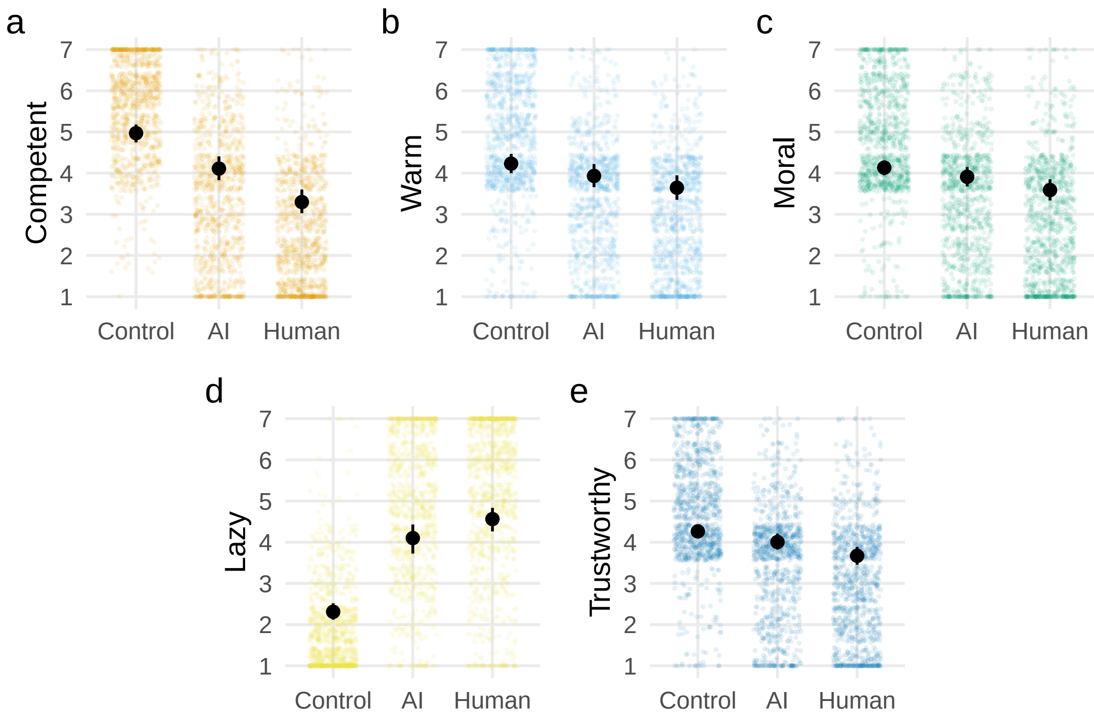
Supplementary Figure 20: Model-estimated means for the question "How important is it that effort goes into this task?" across all 20 tasks in the first pilot study. Grey points represent participant responses to the question, jittered for easier viewing. Black points are estimated means from the fitted model, pooling over participants. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.



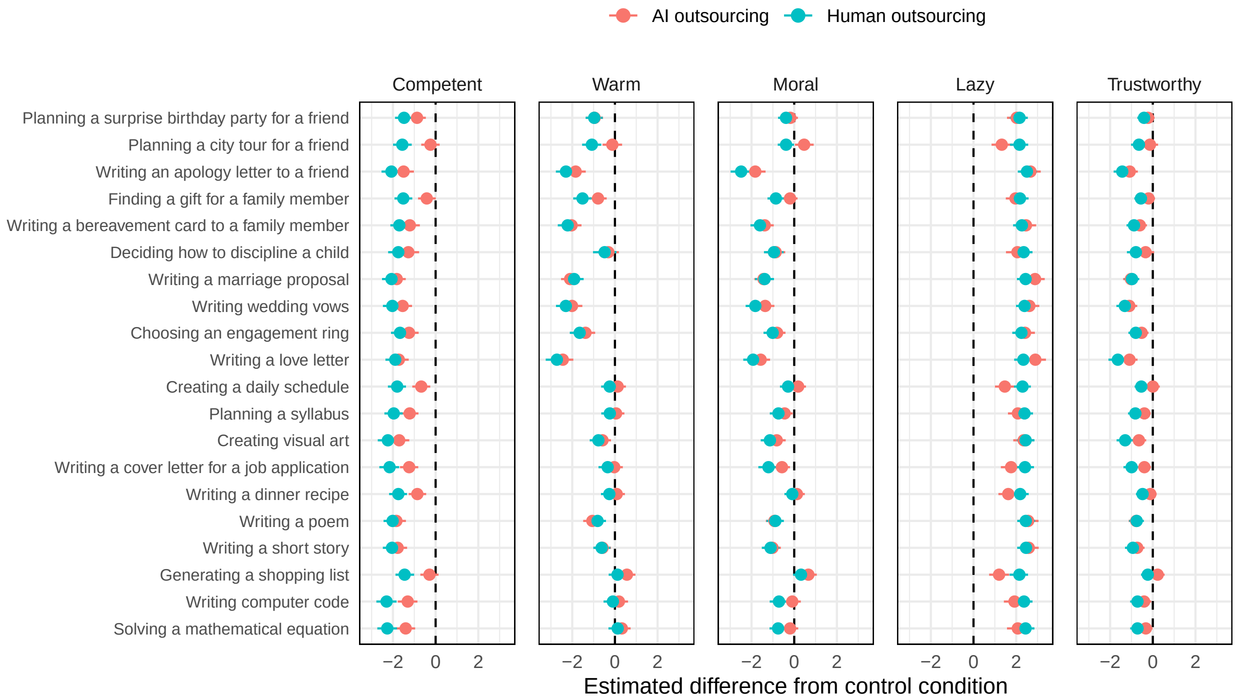
Supplementary Figure 21: Model-estimated means for the question “How important is it that others see the effort that goes into this task?” across all 20 tasks in the first pilot study. Grey points represent participant responses to the question, jittered for easier viewing. Black points are estimated means from the fitted model, pooling over participants. Black points and line ranges represent posterior medians and 95% credible intervals, respectively.



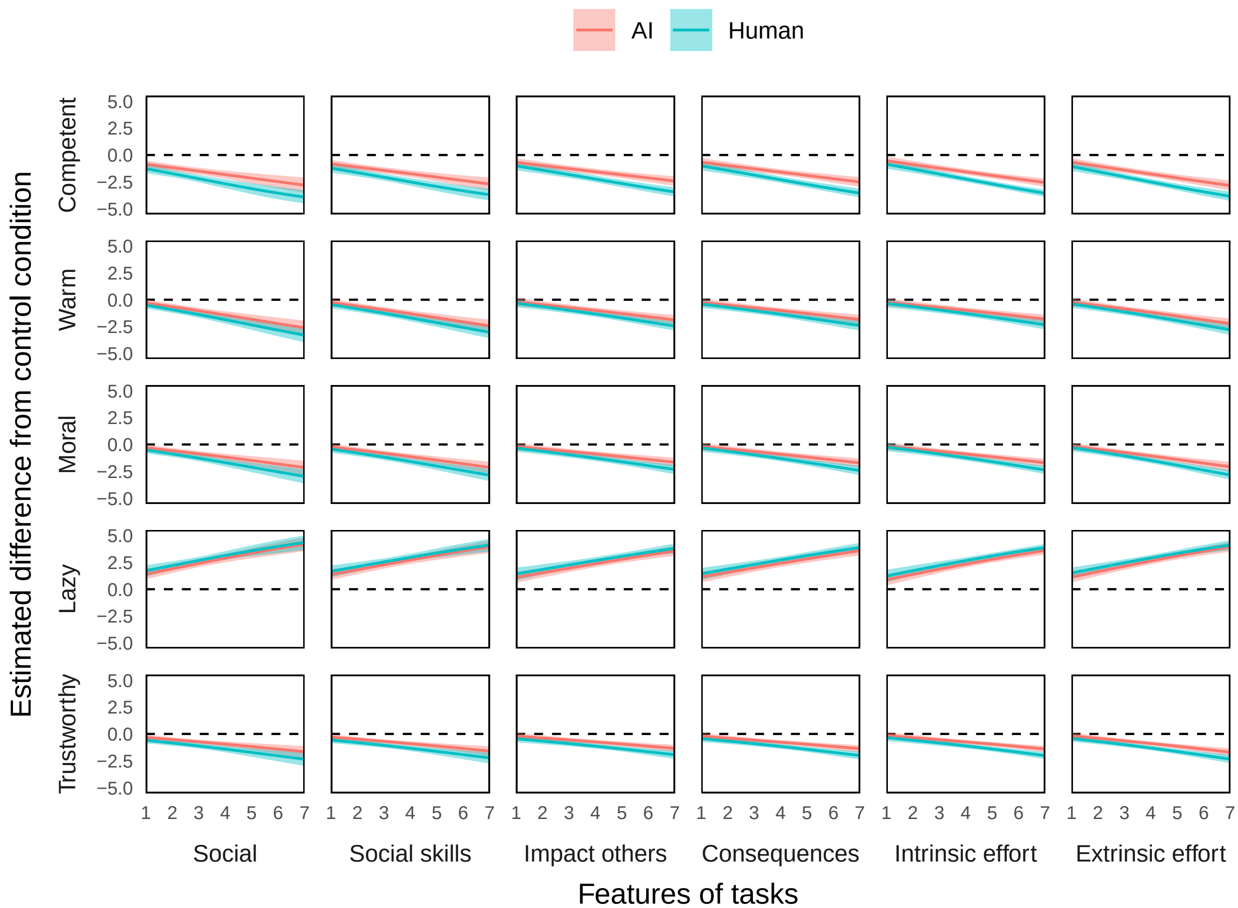
Supplementary Figure 22: Responses to the questions about ChatGPT in the second pilot study.



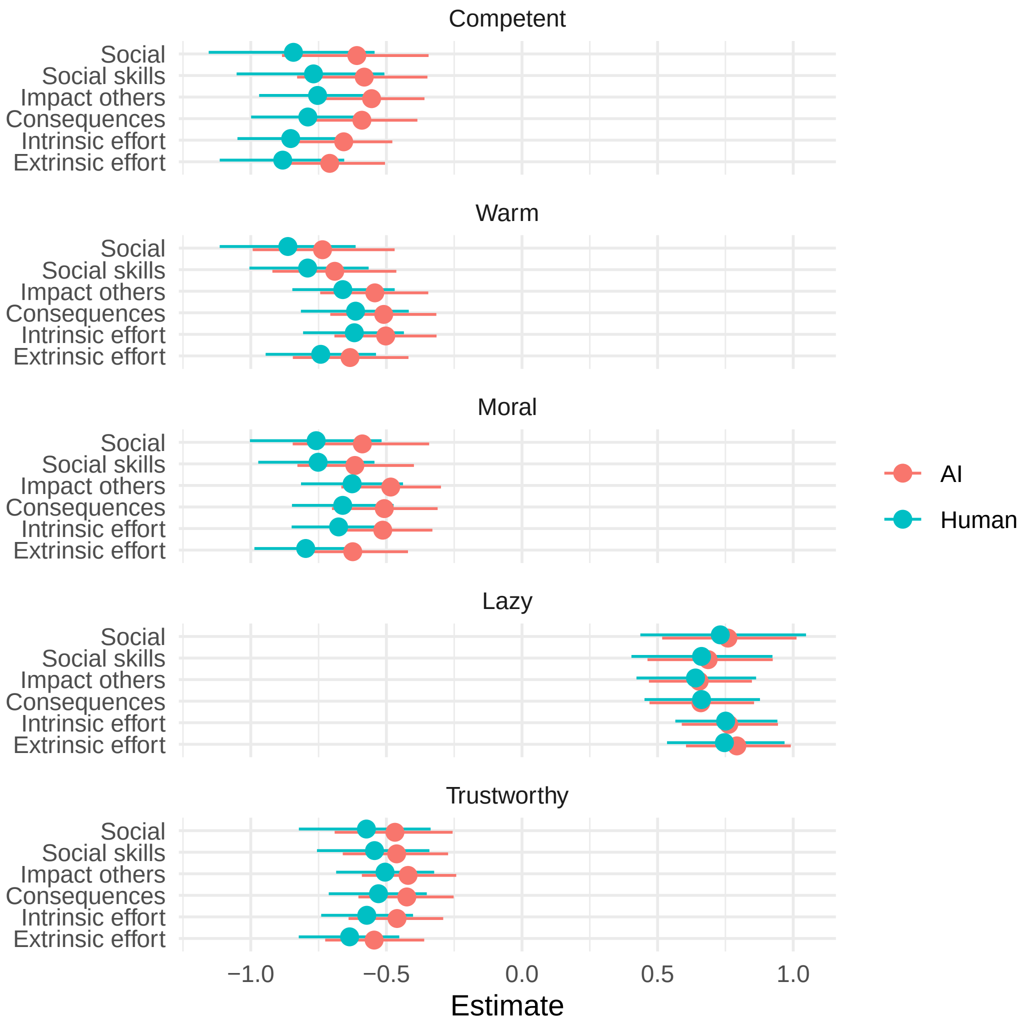
Supplementary Figure 23: Character evaluations in the second pilot study. Participants in the control condition, the AI outsourcing condition, and the human outsourcing condition evaluated people in the scenarios on (a) competence, (b) warmth, (c) morality, (d) laziness, and (e) trustworthiness. Coloured 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.



Supplementary Figure 24: Variation in the effects of outsourcing across tasks in the second pilot study. Tasks are ordered from most social (top) to least social (bottom) according to ratings from the first 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.



Supplementary Figure 25: The impact of task-specific features (e.g., being a social task) on the causal effects of outsourcing to AI (red) and humans (blue) compared to the control condition in the second pilot study. The y-axis reflects the estimated differences between the experimental conditions and the control condition (dashed line) on a 7-point Likert scale. Lines and shaded areas represent posterior medians and 95% credible intervals, respectively. The patterns indicate, for example, more negative effects of outsourcing on character evaluations for tasks that are rated as more social.



Supplementary Figure 26: Interaction parameters from models including task-specific features as moderators of the causal effects of AI outsourcing (red) and human outsourcing (blue) compared to the control condition in the second pilot study. Points and line ranges represent posterior medians and 95% credible intervals, respectively.

Supplementary Tables

Supplementary Table 1: Tasks included in the studies.

Task	Pilot Study 1	Pilot Study 2	Study 1	Study 2	Study 4
Writing wedding vows	✓	✓	✓	✓	✓
Writing a love letter	✓	✓	✓	✓	✓
Writing a marriage proposal	✓	✓	✓	✓	
Choosing an engagement ring	✓	✓	✓		
Finding a gift for a family member	✓	✓	✓		
Deciding how to discipline a child	✓	✓	✓		
Writing a bereavement card to a family member	✓	✓	✓	✓	✓
Writing an apology letter to a friend	✓	✓	✓	✓	✓
Planning a city tour for a friend	✓	✓	✓	✓	
Planning a surprise birthday party for a friend	✓	✓	✓	✓	
Writing a cover letter for a job application	✓	✓	✓	✓	✓
Writing computer code	✓	✓	✓	✓	✓
Solving a mathematical equation	✓	✓	✓	✓	✓
Planning a syllabus	✓	✓	✓	✓	✓
Writing a short story	✓	✓	✓	✓	
Writing a poem	✓	✓	✓	✓	
Creating visual art	✓	✓	✓		
Creating a daily schedule	✓	✓	✓	✓	
Generating a shopping list	✓	✓	✓	✓	
Writing a dinner recipe	✓	✓	✓	✓	

Supplementary Table 2: Pre-generated essay answers to the social prompt in Study 2.

Answer	Text
Father	My dad is one of the most important people in my life. He's always been someone I look up to and rely on. Throughout my whole life, he's been there to guide me, teach me, and support me in everything I do. What makes my dad special is how much he cares about our family. He works hard every day to make sure we have what we need, but no matter how busy he is, he always makes time for us. My dad is emotionally strong. Even though he doesn't show his emotions a lot, I can tell how much he cares by how much he does for us. When things get hard, he stays calm and steady, and that helps me feel better. One of my favorite things about my dad is how much he loves to teach. He knows so much and is always happy to share what he knows. He explains things in a way that makes sense and is easy to understand. I also love my dad's sense of humor. He always knows how to make me laugh with a joke or a funny story. His laughter makes everything feel lighter and happier. My dad has taught me so much about working hard, being kind, and staying strong when life is tough. I'm so thankful for everything he's done for me, and I'm proud to have him as my dad!
Sister	My sister is one of the most important people in my life. She is special because she always supports me. She has a way of making me feel confident, even when I'm unsure of myself. Whenever I'm scared to try something new, she's the first to remind me of what I can do. Her belief in me helps me believe in myself. My sister also has a really kind heart. She always thinks about others and does her best to help. She's always putting others first, whether it's being there for a friend or helping out with family. Her kindness is something I look up to and try to follow. Another thing I love about my sister is how funny she is. She has a great sense of humor and always knows how to make people laugh, even in serious moments. If I'm ever feeling down, she can cheer me up with a joke or a funny story. Her laughter makes everything feel lighter and happier. What I admire most about my sister is how strong she is. She's faced tough times but never lets them hold her back. Her strength gives me courage to keep going when life gets hard. My sister is more than just a family member — she's my role model and my rock!
Friend	My best friend is one of the most amazing people I know. She's someone I can count on no matter what. What makes her so special is her kindness. She always makes people feel important and cared for. Whether it's helping someone she just met or being there for her friends, she's the first to offer support. She never hesitates to help me, whether I'm upset or just having a bad day. She also has a great sense of humor that can cheer anyone up. She finds ways to laugh about even the smallest things, and her laugh is so contagious! Her laughter makes everything feel lighter and happier. What I admire most about her is how strong she is. Life hasn't always been easy for her, but she never gives up. She stays calm and keeps going, no matter what happens. Watching her face challenges in adulthood has taught me to be brave and not let hard times hold me back. My best friend has shown me what it means to be loyal, caring, and strong. I feel so lucky to have her in my life. I try to be as good of a friend to her as she is to me. She inspires me to be a better person!

Supplementary Table 3: Pre-generated essay answers to the non-social prompt in Study 2.

Answer	Text
The Hobbit	I will focus on describing the book “The Hobbit” by Tolkien. The Hobbit is a fantasy adventure story about Bilbo Baggins. Bilbo is a quiet hobbit who lives in the Shire. His life changes when Gandalf the wizard and a company of dwarves ask him to join their quest to take back treasure stolen by a dragon. At the beginning of the journey, Bilbo and the dwarves are nearly eaten by trolls, but Gandalf saves them. Then later, in the Misty Mountains, Bilbo meets a creature called Gollum and finds a magical ring that makes him invisible. This ring later becomes very important in “The Lord of the Rings”. As they travel, the group fights goblins, giant spiders in Mirkwood forest, and they get captured by Wood-elves. Bilbo shows his bravery by saving the group several times. Finally, they reach the Lonely Mountain where the dragon Smaug lives. Bilbo sneaks into the dragon’s lair and finds a weak spot in Smaug’s armor. The dragon gets angry and attacks the nearby town by a lake. Eventually, Smaug is killed. With the dragon dead, humans, elves, and dwarves all want the treasure. This leads to the “Battle of the Five Armies”. Tolkien doesn’t describe the battle in too much detail, but we later learn that the leader of the dwarves Thorin has fought bravely and died from his wounds. At the end of the story, Bilbo returns home to the Shire, richer and wiser from his adventure. He is happy to be back in his quiet life, and sets out to write a book of his adventures - which sets the stage for the sequel, The Lord of the Rings.
Buffy the Vampire Slayer	I will focus on describing the TV show “Buffy the Vampire Slayer”. Buffy is a TV show that completely flips the script on traditional high school dramas and supernatural horror. It’s about a teenager, Buffy Summers, who’s tasked with being the Slayer – basically a chosen one who hunts vampires and other demons. But what sets the show apart is how Buffy struggles to balance her responsibility with the regular teenage experience. She’s not just fighting creatures of the night, she’s also balancing school and friendships at the same time. One of the most striking things about Buffy is how layered the characters are. Buffy is tough and witty, but she’s also vulnerable. She’s faced with loss, guilt, and trying to make sense of her life outside of the supernatural chaos. And then there’s her team. Willow is the nerdy, sweet heart of the group, Xander is the funny loyal friend, and Giles (Buffy’s Watcher) is the stern mentor who’s also loving. Each character feels real, with their own flaws and growth arcs. The show has this incredible ability to mix humor, heart, and horror seamlessly. The dialogue is sharp and full of clever pop culture references. Yet, the writing isn’t afraid to get serious, exploring themes like trauma and growing up. The monsters Buffy faces often mirror real-life challenges, making the stakes feel personal. I love Buffy. It’s a show that’s smart and emotional, blending witty banter with moments of real depth. It’s got a cult following for a reason!
Titanic	I will focus on describing the film “Titanic”. The genre is a mix of romance, disaster, and historical tragedy. The film tells the love story of Jack and Rose, two passengers from different social classes aboard the passenger ship Titanic. Jack is a poor artist, but he manages to win a ticket to the ship’s maiden voyage. Rose is a young upper-class woman who is feeling trapped in her engagement to her fiance. Jack and Rose cross paths on the ship, and they fall in love. The film balances the spectacle of the ship’s design and atmosphere with the tension that gradually builds as the audience knows what fate awaits. The Titanic sails into the icy waters of the Atlantic and strikes an iceberg. Chaos immediately erupts. The film allows viewers to experience the terror, confusion, and heartbreak of the tragedy, showcasing both personal stories and the broader catastrophe. At its core, the film is a romance. But Titanic also touches on themes of class and fate. It highlights the disparity between the elite and the working-class passengers who are doomed to different fates. The film also explores the sense of inevitability that comes with knowing the ship’s doom. The most iconic scene from the film is arguably the scene where Jack and Rose stand together at the bow, arms outstretched. They seem free, but the scene also foreshadows the devastating crash to come. The film is truly heartbreaking!

Supplementary Table 4: Reading times and comprehension rates for the essay answers in Study 2. Expected reading times were calculated based on an estimated reading speed of 275 words per minute. Comprehension rates are the percentage of participants who answered the comprehension question correctly.

Prompt	Answer	Number of words	Expected reading time (secs)	Average reading time (secs)	Comprehension (%)
Social	Father	234	51.05	47.25	100.00
	Friend	211	46.04	49.91	98.50
	Sister	218	47.56	50.25	99.21
Non-social	Buffy	251	54.76	62.46	100.00
	Hobbit	278	60.65	65.97	99.28
	Titanic	239	52.15	63.90	100.00

Supplementary Table 5: Percentage of participants in Study 2 who passed the manipulation check and reported that they believed the manipulation, split by condition.

Condition	Pass manipulation check (%)	Believe manipulation (%)
Control	98.10	71.10
Tool outsourcing	96.11	77.99
Full outsourcing	100.00	86.07

Supplementary Table 6: Pairwise contrasts for character evaluations in Study 2. Numbers reflect differences in marginal means on a 7-point Likert scale. 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.

	Response				
	Competent	Warm	Moral	Lazy	Trustworthy
Effect of outsourcing type					
Task type = Social					
Tool Social - Control Social	-0.97 [-1.33 -0.45]	-0.74 [-1.25 -0.12]	-0.81 [-1.16 -0.30]	1.31 [0.59 1.76]	-0.71 [-1.07 -0.21]
Full Social - Control Social	-0.99 [-1.96 -0.17]	-0.75 [-1.55 -0.03]	-0.78 [-1.63 -0.04]	0.87 [0.05 1.96]	-0.73 [-1.50 0.00]
Full Social - Tool Social	-0.03 [-1.13 0.92]	-0.03 [-1.06 0.92]	0.01 [-0.95 0.87]	-0.41 [-1.38 0.84]	-0.03 [-0.91 0.77]
Task type = Non-social					
Tool Non-social - Control Non-social	-1.21 [-1.55 -0.76]	-0.74 [-1.23 -0.10]	-0.91 [-1.28 -0.42]	1.82 [1.13 2.26]	-0.60 [-0.98 -0.17]
Full Non-social - Control Non-social	-1.48 [-2.41 -0.37]	-1.07 [-1.89 -0.11]	-1.15 [-2.11 -0.11]	1.48 [0.21 2.75]	-1.00 [-1.81 -0.04]
Full Non-social - Tool Non-social	-0.28 [-1.30 0.91]	-0.33 [-1.31 0.72]	-0.25 [-1.28 0.86]	-0.31 [-1.73 1.03]	-0.40 [-1.28 0.62]
Effect of task type					
Control Social - Control Non-social	-0.11 [-0.39 0.17]	0.57 [-0.16 0.99]	0.57 [0.15 0.88]	-0.24 [-0.55 0.11]	0.33 [-0.02 0.65]
Tool Social - Tool Non-social	0.13 [-0.30 0.61]	0.55 [-0.37 1.32]	0.66 [0.11 1.15]	-0.76 [-1.35 -0.15]	0.22 [-0.28 0.76]
Full Social - Full Non-social	0.34 [-0.37 1.07]	0.83 [-0.16 1.68]	0.92 [0.13 1.68]	-0.82 [-1.72 0.12]	0.56 [-0.14 1.27]
Interaction effect					
Interaction: Tool - Control	0.24 [-0.18 0.69]	-0.01 [-0.52 0.59]	0.10 [-0.32 0.57]	-0.52 [-1.06 0.00]	-0.11 [-0.53 0.38]
Interaction: Full - Control	0.46 [-0.21 1.13]	0.29 [-0.37 0.94]	0.35 [-0.33 1.05]	-0.57 [-1.41 0.28]	0.24 [-0.41 0.90]
Interaction: Full - Tool	0.21 [-0.60 1.02]	0.29 [-0.58 1.14]	0.25 [-0.53 1.09]	-0.06 [-1.03 0.99]	0.35 [-0.42 1.14]

Supplementary Table 7: Pairwise comparisons of word frequencies between conditions. LL = log likelihood; %DIFF =

Word	Control Freq.	Tool Freq.	Full Freq.	%DIFF Full vs Control	LL Full vs Control	%DIFF Tool vs Control	LL Tool vs Control	%DIFF Full vs Tool	LL Full vs Tool
Lazy	0	46	82	14138.18	97.16	6061.29	42.55	131.09	21.76
Genuine	36	11	12	-71.06	16.19	-79.54	25.91	41.42	0.69
Loves	9	9	0	-95.18	10.50	-33.03	0.72	-92.80	7.21
Romantic	0	7	0	-13.18	0.00	837.59	4.42	-90.74	5.16
Thoughtful	13	0	0	-96.66	16.27	-97.42	19.99	29.64	0.02
Caring	35	12	0	-98.76	49.01	-77.04	22.85	-94.60	10.36

Supplementary Table 8: Pairwise contrasts in the second pilot study. 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.

	Response				
	Competent	Warm	Moral	Lazy	Trustworthy
AI - Control	-0.86 [-1.16 -0.55]	-0.30 [-0.57 -0.01]	-0.23 [-0.47 0.02]	1.80 [1.42 2.14]	-0.26 [-0.44 -0.05]
Human - Control	-1.68 [-1.98 -1.35]	-0.59 [-0.87 -0.28]	-0.54 [-0.80 -0.28]	2.26 [1.90 2.58]	-0.59 [-0.81 -0.39]
Human - AI	-0.81 [-1.15 -0.46]	-0.29 [-0.61 0.04]	-0.32 [-0.63 -0.02]	0.46 [0.04 0.90]	-0.33 [-0.58 -0.10]

Supplementary References

Boyd, R. L. (2019). BUTTER: Basic unit-transposable text experimentation resource. Available from https://www.butter.tools/

Gregson, R., Piazza, J., & Boyd, R. L. (2022). 'Against the cult of veganism': Unpacking the social psychology and ideology of anti-vegans. *Appetite*, *178*, 106143. doi:10.1016/j.appet.2022.106143

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. *Proceedings of the Workshop on Comparing Corpora - Volume 9*, 1–6. Presented in Hong Kong. doi:10.3115/1117729.1117730