

Social Psychology in the Age of AI
EASP 2026 Preconference
Organised by Mengchen Dong, Marcos Dono, and Jim A.C. Everett

Schedule at a Glance

Time	Format	Content
09.00 – 10.15	Talk Panel 1 Four 15 minute talks; each followed by short Q&A	A Small Misstep for Man, a Giant Mistake for AI: Variance Beliefs and Human–AI Impressions Schulman et al. Social Perceptions of Deadbots: Appropriateness, Risks, and Conditional Acceptance of Posthumous AI Riva et al. Choosing Empathy in the Age of AI: Psychological and Normative Considerations Cameron et al. Substratism: Conceptualizing and Measuring Moral Bias Against AI Ladak et al.
10.15 – 10.30	Coffee Break	
10.30 – 11.20	Blitz Panel 1 Five 5-minute talks + 25 minute group discussion	The Role of Egocentrism in Moral and Fairness Judgments of AI and Human Decisions Miazek et al. What Moral Motivations Do People Attribute to AI—and Which Ones Matter for Trust? A Multi-Country Study McKee et al. Why Do People Punish AI? de Mello et al. Gemini Is for Liberals and Grok for Conservatives? The Effect of Ideological Similarity on Trust in LLMs Wingen et al. People View Delegating All Research Tasks to GenAI as Untrustworthy, Inaccurate, and Immoral Conway
11.20 – 11.30	Short Break	
11.30 – 12.30	Roundtable 1	“Under the scope of various social realities, does AI bring

		more benefit or harm?"
12.30 – 13.30	Lunch	
13.30 – 14.45	Talk Panel 1 Four 15 minute talks; each followed by short Q&A	The State of AI Use in Academic Research and its Unintended Consequences Bindal et al. For Youth, by Youth: Co-producing Safeguards for Generative AI Use with UK Adolescents Reinecke et al. Human-in-the-Loop Oversight of AI is Compromised by Political Preferences Dores-Cruz et al. Perceived Intergroup Competition Shapes Public Consent for Deregulating Artificial Intelligence Veitch et al.
14.45 – 15.00	Coffee Break	
15.00 – 15.50	Blitz Panel 1 Five 5-minute talks + 25 minute group discussion	LLM vs. Human Affective Connotations of Identities, Behaviours, and Settings Across Three Cultures Blaison et al. Supporting Non-Traditional Students in Distance Education with AI-Based Interventions Froehlich et al. An AI Answer is Worth 1/5 of a Survey Response: The Disconnect Between Explicitly Reported Trust and Empirical Belief Updating Teitelbaum et al. Suspicious of AI? Perceived Autonomy and Interdependence Predict AI-Related Conspiracy Beliefs Qi et al. The Existence of Manual Mode Increases Human Blame for AI Mistakes Meyers et al.

15.50 – 16.00	Short Break	
16.00 – 17.00	Roundtable 2	“How can social psychology inform AI development and regulation?”

Full Programme and Abstracts

09:00-10:15

Talk Panel 1

A Small Misstep for Man, a Giant Mistake for AI: Variance Beliefs and Human–AI Impressions

Roy Schulman, Yochanan Bigman (Business School, The Hebrew University); Melissa Ferguson, (Psychology Dept., Yale University)

As AI increasingly functions as a social agent in everyday life, how people form and updated impressions of AI agents becomes increasingly important. From a theoretical perspective, it is important to understand whether, and how, people update their impressions of AIs differently than their impressions of humans. Classic models on impression updating suggest that new information is weighted by its diagnosticity, which depends in part on perceived behavioral variability. However, beliefs about an agent’s natural performance variability have rarely been measured or manipulated directly. In six studies (N = 184; 292; 290; 572; 851; 555) we find that people believe the performance of AI has less variance than the performance of humans, leading people to update their impressions of AI faster than their impressions of humans. Studies 1a–1b show that people systematically expect AI agents to exhibit less performance variability than humans across diverse tasks. Studies 2–4 demonstrate that identical performance changes—whether improvements or declines—lead to stronger updating for AI than for humans. Study 5 experimentally manipulates perceived variability and shows that equating variance information eliminates this updating asymmetry. These findings identify variance beliefs as a key mechanism in impression updating, and underlying differences in how people evaluate humans and AIs, explaining heightened sensitivity to AI errors and trust in the technological improvement of AI.

Social Perceptions of Deadbots: Appropriateness, Risks, and Conditional Acceptance of Posthumous AI

Paolo Riva, Illenia Gasparini, Alessia Telari, Luca Pancani (University of Milano-Bicocca)

Artificial intelligence has enabled applications that simulate interactions with deceased individuals, commonly referred to as deadbots. Despite growing public attention, limited empirical work has examined how such technologies are socially perceived. Across four studies, we investigated perceived appropriateness, risks, and potential of deadbots in the context of digital commemoration. Study 1 (N = 100) used a qualitative survey to examine spontaneous perceptions of deadbots, identifying themes related to emotional reactions, ethical concerns, perceived benefits, and contextual moderators. Studies 2a (N = 170) and 2b (N = 102) employed quantitative measures to assess perceived appropriateness, risks, social norms, and potential benefits, and their associations with individual differences such as attachment styles, personal innovativeness, perceived AI threat, religiosity, and political orientation. Study 3 (N = 137) compared deadbots with traditional digital traces of the deceased, while Study 4 (N = 222) contrasted deadbots with AI-generated deepfake videos and different types of digital traces. Across studies, deadbots were perceived as highly risky and only moderately appropriate. Evaluations were context-dependent: short-term, occasional, curiosity-driven, or consented use was judged as more acceptable than prolonged, grief-driven, or non-consensual use. Individual differences, particularly personal innovativeness, were associated with more favorable evaluations.

Choosing Empathy in the Age of AI: Psychological and Normative Considerations

C. Daryl Cameron, Joshua Wenger (Penn State University), Madeline G. Reinecke (University of Oxford)

Empathically resonating with others often entails motivated choices to regulate emotions in particular ways – a willingness, rather than ability, to engage. In this talk, I will first outline a motivated empathy framework and how it can be used to understand how we relate to our own feelings of empathy. I will discuss how this approach can be applied to human-AI interactions, first by summarizing past work on choices to receive empathy from large language models, and second by summarizing the kinds of questions that relational ethics suggest for these interactions. In the second half of the talk, I will discuss how interactions between humans and AI (e.g., robots, LLM's) can challenge our understanding of what empathy can and should be, and by suggesting that we reconsider benchmarks for what count as empathy as a psychological construct.

Substratism: Conceptualizing and Measuring Moral Bias Against AI

Ali Ladak, Matti Wilks, Steve Loughnan (University of Edinburgh), Jacy Anthis (University of Chicago), Janet Pauketat (Sentience Institute)

Artificial intelligences (AIs) are increasingly taking on social roles, such as virtual coworkers and personal companions. Yet people grant AIs very little moral concern, even when they are described as indistinguishable from humans. We argue that this bias may be substratism: the moral devaluation of AIs due to their non-biological nature (e.g., silicon and wires rather than flesh and blood). Across five preregistered studies (N = 2,129), we introduce substratism as a psychological construct and develop and validate a scale to measure it. In Study 1, we found that substratism is best captured by a unidimensional factor structure and we reduced a large initial item pool into an eight-item scale. Studies 2 and 3 confirmed the factor structure with new samples and showed that substratism is correlated with other AI-related measures, such as perceived threat from AI. However, it was weakly correlated or uncorrelated with other prejudices and their underlying causes, suggesting that substratism has unique psychological explanations. Studies 4 and 5 showed that substratism predicts relevant outcomes: prioritizing humans and animals over AIs in moral decisions, and choosing to learn about an AI rights charity and its petition. Overall, our findings suggest that substratism is a distinct, measurable construct that varies meaningfully across individuals, and we provide a concise, validated scale to capture and quantify this bias against increasingly advanced AI systems.

10:30-11:20

Blitz Panel 1

The Role of Egocentrism in Moral and Fairness Judgments of AI and Human Decisions

Katarzyna Miazek, Konrad Bocian (SWPS University)

Algorithmic fairness is a core principle of trustworthy Artificial Intelligence (AI), yet how people perceive fairness in AI decision-making remains understudied. Prior research suggests that moral and fairness judgments are egocentrically biased, favoring self-interested outcomes. We examine whether this bias extends to AI decision-makers, comparing fairness and morality perceptions of AI and human agents. Across three experiments (two preregistered, N = 1880, Prolific US samples), participants evaluated financial decisions made by AI or human agents. Self-interest was manipulated by assigning participants to conditions where they either benefited from, were harmed by, or remained neutral to the decision outcome. Results showed that self-interest significantly biased fairness judgments—decision-makers who made unfair but personally beneficial decisions were perceived as more moral and fairer than those whose decisions benefited others (Studies 1 & 2) or those who made fair but personally costly decisions (Study 3). However, this egocentric bias was weaker for AI

than for humans, mediated by a lower perceived mind and reduced liking for AI (Studies 2 & 3). These findings suggest that fairness judgments of AI are not immune to egocentric biases, but are moderated by cognitive and social perceptions of AI versus humans. Our studies challenge the assumption that algorithmic fairness alone is sufficient for achieving fair outcomes.

What Moral Motivations Do People Attribute to AI—and Which Ones Matter for Trust? A Multi-Country Study

Paul McKee, Walter Sinnott-Armstrong, Breanna Nguyen (Duke University, USA), Paulo Boggio (Mackenzie Presbyterian University, Brazil), Jim Everett (University of Kent, UK)

AI systems increasingly inform decisions with moral consequences, yet little is known about what moral motivations people believe AI can possess or which are required for trust. Drawing on the Motivation of Moral Action (MoMA) model, which distinguishes four motivations—Approach Good, Avoid Good, Approach Bad, and Avoid Bad—we examine how people across cultures infer AI's moral capacities and how these inferences shape trust. We report one multi-country online survey experiment (pilot N = 700), with approximately 100 participants each from Brazil, India, Japan, Kenya, the United Kingdom, the United States, and Ukraine. Participants evaluate realistic AI systems (e.g., medical triage, financial advising, hiring) and rate which motivations AI can possess, which are required for trust, and their willingness to rely on the system. We test five preregistered hypotheses. First, participants will judge all four motivations as possible for AI systems to possess. Second, prosocial motivations (Approach Good and Avoid Bad) will be viewed as most important for trust, whereas antisocial motivations will undermine trust. Third, inferred motivations will vary across countries. Fourth, cross-national differences in trust will be mediated by attributed motivations. Fifth, cultural and institutional factors will moderate these relationships. Together, these tests clarify how moral expectations of AI vary globally and which motivations are essential for trustworthy and ethical AI.

Why Do People Punish AI?

Victoria Oldemburgo de Mello, Michael Inzlicht (University of Toronto)

People blame and punish AI for wrongdoing—but what are they trying to achieve? Punishment typically serves retribution, deterrence, or reform, yet AI systems cannot suffer, be deterred, or be reformed in the ways humans can. Across three experiments, we investigate the psychological motivations underlying AI punishment. Study 1 shows that when reasoning abstractly about AI wrongdoing, people's first intuition is to sanction creators rather than the AI itself. However, Studies 2 and 3 reveal that when people

directly interact with an AI that committed a moral violation, they choose retributive punishment (removing tokens from the AI in a trust game) over reform and deterrence options—even when they know the AI will not be aware of the punishment. These findings suggest that AI punishment is driven by the same retributive motivations underlying human-human punishment, rather than by instrumental goals that AI punishment could actually achieve. The findings also support the notion of a "responsibility gap" where people's intuitive responses to AI wrongdoing may be psychologically satisfying but functionally ineffective.

Gemini Is for Liberals and Grok for Conservatives? The Effect of Ideological Similarity on Trust in LLMs

Tobias Wingen (FernUniversität in Hagen), Marlene S. Altenmüller (ZPID / University of Trier), Angela R. Dorrough (FernUniversität in Hagen)

Recent work suggests that ideological similarity is a core driver of trust among humans. Building on this perspective, we asked whether trust in Large Language Models (LLMs) may be shaped by a similar process. Specifically, we assumed that when deciding whether to trust an LLM, users perceive the model's political ideology and compare it to their own ideology. The resulting perceived similarity then influences their trust (e.g., liberals trust liberal LLMs, conservatives trust conservative LLMs). We report four studies conducted in the United States and Germany (total N = 1,752). Studies 1 - 3 used U.S.-based online samples recruited via Prolific with quotas for age, gender, and political ideology. Study 1 (N = 499, correlational) showed that ideological similarity with LLMs was positively associated with trust. Study 2 (N = 493, experimental) manipulated the LLM's supposed ideology and demonstrated a causal effect of ideological fit on trust. Study 3 (N = 499) experimentally varied perceived ideological similarity directly and found that similarity increased trust and participants' willingness to rely on the model's recommendations. Study 4 (N = 261), conducted with a convenience sample of German university students, replicated the key findings of Study 3 in a new cultural and political context. Across all studies, the perceived political ideology of LLMs and the resulting ideological similarity thus indeed emerged as central drivers of trust judgments.

People View Delegating All Research Tasks to GenAI as Untrustworthy, Inaccurate, and Immoral

Paul Conway (University of Southampton)

The introduction of ChatGPT has fuelled a public debate on the appropriateness of using Generative AI (large language models; LLMs) in work, including a debate on how they might be used (and abused) by researchers. In the current work, we test whether

delegating parts of the research process to LLMs leads people to distrust researchers and devalues their scientific work. Participants (N = 402) considered a researcher who delegates elements of the research process to a PhD student or LLM and rated three aspects of such delegation. Firstly, they rated whether it is morally appropriate to do so. Secondly, they judged whether—after deciding to delegate the research process—they would trust the scientist (that decided to delegate) to oversee future projects. Thirdly, they rated the expected accuracy and quality of the output from the delegated research process. Our results show that people judged delegating to an LLM as less morally acceptable than delegating to a human ($d = -0.78$). Delegation to an LLM also decreased trust to oversee future research projects ($d = -0.80$), and people thought the results would be less accurate and of lower quality ($d = -0.85$). We discuss how this devaluation might transfer into the underreporting of Generative AI use.

12:30-13:30

Talk Panel 2

The State of AI Use in Academic Research and its Unintended Consequences

Zerevan Bindal, Frances Chen (University of British Columbia, Canada), Stephen Hay (Griffith University, Australia), Toni Schmader (University of British Columbia, Canada)

Artificial Intelligence (AI) tools are transforming the research landscape. While researchers can leverage these tools to make workflows more efficient (Li et al., 2025), they also risk exacerbating existing disparities in research productivity (Tang et al., 2025). In this talk, we present data from N = 1,162 academics engaged in research across 17 disciplines and 42 countries, collected in Fall 2025, to characterize the current state of AI use in academia. We measured AI use frequency (1 = never; 7 = daily) and attitudes toward AI (1 = strongly disagree; 7 = strongly agree) and compare AI use and attitudes across Europe and the rest of the world. We observe higher AI use frequency and more positive attitudes toward AI use for research in developing economies vs. advanced economies, as well as in men vs. women. Based on these patterns, we discuss the potential for AI use to both narrow and exacerbate existing economic and gender inequities in academia. Our data highlight positive pathways for AI to narrow productivity gaps by increasing academics' self-efficacy in AI use for research-related tasks, especially in developing economies. However, our data also highlight disparities in users' concerns about ethical use, reliability, and lack of regulation that could exacerbate existing gender inequities in academia. Together, these insights highlight unintended consequences of AI use in science and inform a more equitable and responsible integration of AI in academia.

For Youth, by Youth: Co-producing Safeguards for Generative AI Use with UK Adolescents

Madeline G. Reinecke, Kathryn B. Francis, William Hohnen-Ford, Ilina Singh
(University of Oxford)

Young people are among the most enthusiastic adopters of generative AI, yet they remain one of its most vulnerable user groups. Moreover, they are largely absent from conversations about how these systems should be used safely and ethically. In this work, we present insights from a youth-led, co-production research programme that identifies critical risks in adolescent generative AI use and develops developmentally appropriate safeguards. Supported by the UK Government's AI Security Institute, the project is designed to inform near-term policy on safe generative AI use by youth. In collaboration with the NeurOX Young People's Advisory Group (young people aged 14–25), adolescents acted as co-researchers rather than participants—mapping high-risk AI use contexts (e.g., AI companionship), co-designing candidate safeguards (e.g., mandatory time limits), and testing mitigation strategies in a sandboxed AI environment. From this process, we report an adolescent-specific risk taxonomy, a co-produced “For Youth, By Youth” safety toolkit, and best-practice guidance for developers and policymakers. Beyond these substantive outputs, we note that co-production methodology remains under-recognised within social psychology. The young people involved were engaged as expert collaborators, underscoring the value of incorporating youth expertise into AI ethics and psychological research.

Human-in-the-Loop Oversight of AI is Compromised by Political Preferences

Terence Dores Cruz, Christopher Starke (University of Amsterdam), Tim Katzke, Emmanuel Müller (Technical University of Dortmund), Marta Kwiatkowska (University of Oxford), Orly Lobel (University of San Diego), Nils Köbis (University Duisburg-Essen), and Shaul Shalvi (University of Amsterdam)

Humans are mandated to oversee sensitive Artificial Intelligence (AI) systems, ensuring final decisions are human-approved. Yet, behavioral science shows that human judgment is prone to motivated reasoning. Here, we report the results of three experiments involving over 115,940 decisions (Study 1: $n = 2545$; 50,880 decisions, Study 2: $n = 1605$; 32,100 decisions; Study 3: $n = 1648$, 32,960 decisions) modeled on real-world welfare allocation decisions. The task involved providing financial support to locals and immigrants, where participants acted as human-in-the-loop (HITL) reviewers of algorithmic recommendations with and without systematic bias. Results reveal that HITL oversight often fails to correct algorithmic errors, often even exacerbating them.

Moreover, Democrats and Republicans exhibit distinct partisan biases in their decisions. Interventions, such as incentivizing accuracy or clarifying instructions, yielded only modest improvements, primarily among Democrats. Nevertheless, HITL error rates were not lower than algorithmic error rates. These findings suggest that HITL oversight does not guarantee reductions in AI errors, underscoring the need for a more nuanced, behavioral science-inspired approach to governing AI oversight.

Perceived Intergroup Competition Shapes Public Consent for Deregulating Artificial Intelligence

Pierce Veitch, Scott Claessens, Jim A.C. Everett (University of Kent)

Governments increasingly frame artificial intelligence (AI) development as a geopolitical race, invoking fears of outgroup dominance to justify weakening safety regulation. Scholars have warned against such “AI race” rhetoric, yet there is no empirical work on how this rhetoric shapes public attitudes. Across four studies in the US, UK, and Australia, we examine how intergroup threat and socio-ideological dispositions relate to support for AI deregulation. In Study 1 (N = 913), people reported realistic and symbolic threats from AI. Both outgroup threat and trust in AI predicted endorsement of deregulation, in addition to collective narcissism and social dominance orientation. In Studies 2a–2b (N = 762), threat-based appeals of outgroup AI development increased perceived threat but not explicit support for deregulation. Nonetheless, such appeals increased consent to revoking regulatory policies (Study 3; N = 595), through soft attitudes (e.g. tolerance, acceptance and leader evaluations). These findings reveal how threat-based rhetoric undermines AI regulation.

14:45-15:00

Blitz Panel 2

LLM vs. Human Affective Connotations of Identities, Behaviours, and Settings Across Three Cultures

Christophe Blaison (Université Paris Cité), Aidan Combs (University of Bamberg; The Ohio State University), Diego Dametto (Potsdam University of Applied Sciences, Free University of Berlin), Renee Leung, (The Laboratory of Data Discovery for Health, Hong Kong), Aarti Malhotra (University of Waterloo, Canada), Tobias Schröder (Potsdam University of Applied Sciences) Jesse Hoey (University of Waterloo, Canada), Lynn Smith-Lovin (Duke University, USA)

This study evaluates whether large language models (LLMs) can replicate the Evaluation (E), Potency (P), and Activity (A) dimensions of human affective meaning across different

cultures. Using prompts that mirror traditional semantic differential surveys, we compared ratings from GPT-4o, Mistral Large, and Llama 3.1 against high-quality human datasets. We compared LLM outputs to mean human ratings from 2,615 U.S. participants (2,481 terms), 700 French participants, and 800 German participants (1,139 terms each). Stimuli terms included social identities, behaviors, and settings. LLMs were prompted in the native language to act as an "average citizen" of the target culture, with each concept rated five times per model. LLM ratings correlate strongly with human means, particularly for Evaluation (up to $r=0.95$ for English). However, results reveal a persistent English-language/American bias; models performed best on U.S. English terms, even when developed by a French company (Mistral). Furthermore, LLMs tend to be affectively extreme, producing overly bimodal distributions compared to more nuanced human judgments. While LLMs are promising tools for measurement, these systematic cultural and structural biases must be accounted for in cross-cultural social psychological research.

Supporting Non-Traditional Students in Distance Education with AI-Based Interventions

Laura Froehlich, Nathalie Bick, Marcus Specht, Stella Kolarik, (FernUniversität in Hagen, Germany)

Students in higher education have increasingly diverse backgrounds concerning sociodemographic attributes and prior experience. Non-traditional students (e.g., non-native speakers, students with low SES or chronic illness/disability) face negative competence-related stereotypes, which are likely to be activated in online learning contexts due to scarce individuating information. AI has the potential to adaptively support students by drawing on surveys and behavioral data with machine learning and learning analytics. I will highlight how an algorithmic system could be used to identify at-risk students and how social-psychological interventions aimed at reducing social identity threat or increasing sense of belonging could be adaptively presented to these at-risk students based on an LLM infrastructure. Moreover, LLM-based conversational agents can also be implemented in computer-supported collaborative learning to monitor and manage virtual collaboration. Ethical considerations concerning algorithmic bias in such AI-based interventions will be discussed. Future social-psychological research could investigate how conversational agents are perceived as social interaction partners in individual and collaborative settings and might benefit or hinder learning and performance.

An AI Answer is Worth 1/5 of a Survey Response: The Disconnect Between Explicitly Reported Trust and Empirical Belief Updating

Louis Teitelbaum, Avihay Hajama, Almog Simchon (Ben-Gurion University of the Negev)

AI responses are based on millions of training examples, leading users to treat them as indications of general consensus. In this way, querying an AI on a matter of public opinion is analogous to taking the mean of many human survey responses. But do human users trust AI responses to an extent compatible with the belief that they represent the average of millions of human answers? We use a Bayesian cognitive modeling approach to estimate the implicit weight of evidence people attribute to AI responses—as opposed to human responses—by observing the degree to which they update their opinions after seeing a single such response. In two studies (N=99; 101), we find that both Israeli students and British Prolific workers explicitly report that one ChatGPT response is worth more than a sizable poll (median size=100; 30), but implicitly weight ChatGPT’s answers much lower than those of anonymous individual humans. Analyzing the empirical accuracy of ChatGPT responses, we find that trust in AI is only slightly lower than the theoretical optimum for our task. Thus it may be that trust in AI is low in certain knowledge domains because AI is indeed unreliable in these domains. Planned follow-up studies investigate whether this pattern is unique to normative questions as opposed to factual world knowledge. Our novel task and modeling approach will be of use to AI researchers and social scientists interested in quantifying the persuasive strength of AI across topics, models, and cultures.

Suspicious of AI? Perceived Autonomy and Interdependence Predict AI-Related Conspiracy Beliefs

X

The Existence of Manual Mode Increases Human Blame for AI Mistakes

Samuel Meyers (The Hebrew University of Jerusalem), Mads N. Arnestad (BI Norwegian Business School), Kurt Gray (University of North Carolina at Chapel Hill), Yochanan E. Bigman (The Hebrew University of Jerusalem)

This abstract is based on findings reported in Arnestad et al., 2024.

People are offloading many tasks to artificial intelligence (AI)—including driving, investing decisions, and medical choices—but it is human nature to want to maintain ultimate control (Ashford & Black, 1996; Burger & Cooper, 1979). Therefore, we suggest that people also want the option of a “manual mode” to take control of the AI-powered machine. However, this need for control may have an unintended consequence –

increased blame (Alicke, 2000), due to perceived causation of harm (Malle et al., 2014) and counterfactual thinking (Epstude & Roese, 2008). So, we propose that the mere existence of a manual mode in autonomous machines will increase its owner's blame for mistakes. We find support for our hypotheses in seven studies ($N = 1838$). First, people prefer to buy a self-driving car with a manual mode than without, despite added costs and better AI-driving than humans (87%; $\chi(1) = 53.44$, $p < .001$). Second, across contexts including self-driving cars, surgery, radiology, and financial investing, the human agent with an option to use manual mode (vs without) is blamed more for an AI mistake (Mean Cohen's $d = 0.81$). Third, this effect is explained independently by perceptions of causation ($b = 0.63$, $CI.95[0.47-0.82]$) and counterfactual cognition ($b = 0.52$, $CI.95[0.37-0.67]$). These results highlight a tension between desire for control and blame and its importance given the increasing role of AI in our lives.