UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Responsibility for influencing others

Permalink

https://escholarship.org/uc/item/6qj1w4k9

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Hu, Chuqi
Gerstenberg, Tobias
Wu, Sarah A

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Responsibility for influencing others

Chuqi Hu[1*], **Tobias Gerstenberg**[2], **and Sarah A. Wu**[2]
[1]Symbolic Systems Program, Stanford University
[2]Department of Psychology, Stanford University
[*]chuqihu@stanford.edu

Abstract

Collective outcomes often result from complex social dynamics where individuals both contribute directly and also shape each other's contributions. How do we hold people responsible for an outcome when their actions influence others? Here, we examine how an individual's role within a group (whether they can influence others, be influenced by others, or act independently) affects how responsible they are judged to be. Across three experiments spanning both social and physical settings, we find that people systematically assign greater responsibility to those who can influence others. Furthermore, influencers with knowledge of their potential impact were held more responsible than those who were unaware. The relative responsibility of individuals who were influenced by others and who acted independently differed by context. Together, these results show that we hold others responsible by considering not only how their actions directly affect the outcome, but also how they affect others' propensity to act.

Keywords: responsibility attribution; causal reasoning; social influence; group decision-making

Introduction

On National Voter Registration Day in 2023, Taylor Swift posted an Instagram story that drove over 35,000 new registrations, a 23% increase from the previous year, including more than twice as many registrations among 18-year-olds (Sullivan, 2023). A Newsweek poll conducted in early 2024 found that 18% of all voters were more likely to vote for a candidate she endorsed (Fung, 2024). Swift's outsized influence on American politics is widely acknowledged, despite her own vote carrying the same weight as anyone else's. Collective outcomes such as elections, petitions, markets, and climate change often result from more than a simple aggregation of individual contributions: they are also driven by the ways in which people influence each other's behavior. How do we hold people responsible not only for the direct effects that their actions have on the outcome, but also for their influence on others' actions?

Previous research has identified many factors that influence how we hold people responsible for group outcomes. These include a person's causal role in bringing the outcome about (e.g. Lagnado et al., 2013; Langenhoff et al., 2021; Spellman, 1997), how much the person contributed to the outcome (e.g. Gerstenberg & Lagnado, 2010; Xiang et al., 2023), inferences about the person's character or mental states (e.g. Cushman, 2008; Gerstenberg et al., 2018; Lagnado & Channon, 2008; Malle et al., 2014; Shaver, 1985), and how easily the person could have been replaced (e.g. Gantman et al., 2020; Wu & Gerstenberg, 2024). Prior work has developed computational models of responsibility that formalize an agent's causal role by evaluating how the outcome would have been different in a counterfactual situation in which the agent had acted differently (e.g. Chockler & Halpern, 2004; Halpern & Pearl, 2005; Wu et al., 2023).

Much of this prior research has focused on outcomes in which agents each contribute directly. However, in most real-world situations, people do not act in isolation; they observe, mimic, guide, and persuade each other. Social norms and affiliations are powerful drivers of collective behavior (Allcott, 2011; Asch, 1951; Cialdini et al., 1990; Duffy & Chartrand, 2015; Festinger, 1954), as people may act to conform to others' expectations or accept others' judgments as evidence of reality (Deutsch & Gerard, 1955). Moral exemplars can provide standards for how one ought to behave (van de Ven et al., 2019). Individuals who set bad examples can be blamed for their imitators' wrongdoings (Kardos et al., 2024). Responsibility may even extend beyond direct causal influence; people can be held responsible for others' actions simply by sharing group membership with them (Lickel & Onuki, 2015; Lickel et al., 2003; Sanders et al., 1996; Shultz et al., 1987).

In this paper, we explore how social influence affects responsibility judgments. We designed situations where multiple causes contributed to an outcome and each cause was one of three types: a *root* cause, which has the potential to influence others; a *dependent* cause, which is influenced by a root cause; and an *independent* cause, which neither influences nor is influenced by others. We designed a paradigm to directly pit these causes against each other while holding equal their contribution to the outcome. Experiment 1 featured voting scenarios in which some individuals' votes were visible to others. We additionally manipulated whether or not these voters were aware of their potential influence. Experiment 2 looked at a physical domain in the form of a system of interacting gears, where one gear's motion could affect another's. Experiment 3 tested similar voting scenarios as in Experiment 1, but framed influence as receiving voting advice from another person.

Using these paradigms, we investigated how responsibility is distributed across different roles within a group. We hypothesized that root causes would be held more responsible than independent causes, who would in turn be held more

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

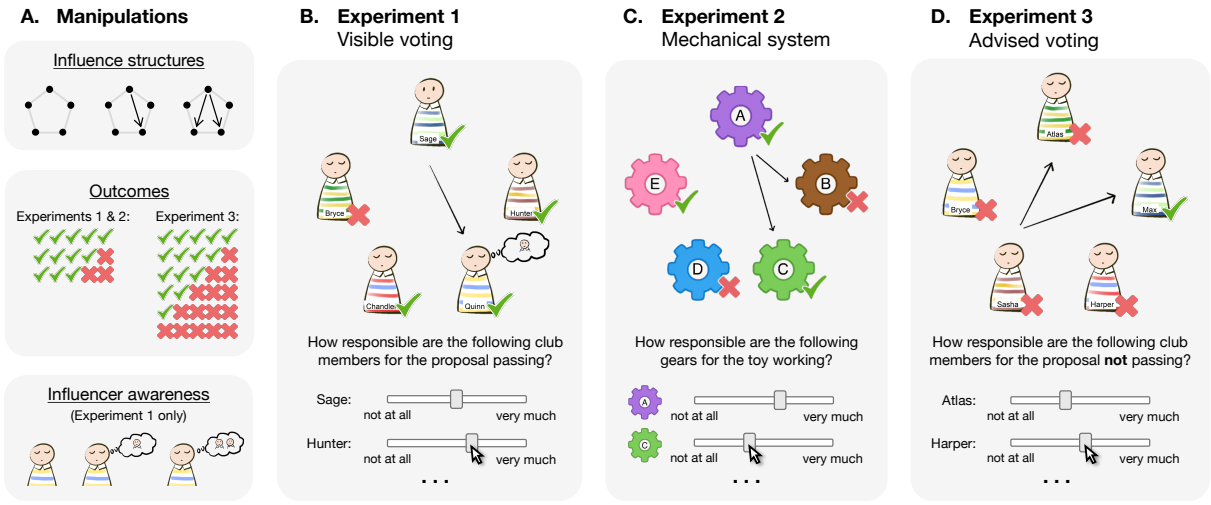


Figure 1: **Experiment overview.** (A) Participants made responsibility judgments for five different contributors in groups where we manipulated the influence structure, the outcome, and additionally whether the root cause was aware of their influence in Experiment 1. (B) In Experiment 1, we designed a voting paradigm and focused on positive outcomes only with the additional awareness manipulation. (C) In Experiment 2, we designed a physical setting and again focused on positive outcomes. (D) In Experiment 3, we framed influence as a mentor advising a mentee, and tested both positive and negative outcomes. Note that the arrows point from dependent causes to the root cause in Experiment 1, and from the root cause to dependent causes in Experiments 2 and 3.

responsible than dependent causes, for several reasons. First, this ordering might reflect the relative agency of the different actors. Agents perceived to have greater capacity to act and control the outcome tend to be held more responsible than those who are "acted upon" (Gray & Wegner, 2009; Malle et al., 2014). Second, dependent causes may be held the least responsible because they lie along a causal chain initiated by a root cause, which discounts the role of their own choice (Engelmann & Waldmann, 2022; Hilton et al., 2010; Kelley, 1973; McClure, 1998). Finally, root causes may be perceived to bring about the greatest change in the probability of the outcome, as their actions include the potential addition of others' (Brewer, 1977; Gerstenberg et al., 2023; Spellman, 1997). We additionally hypothesized that root causes who were aware of their potential influence would be held more responsible than those who were unaware, because they had greater foreseeability over the outcome (Lagnado & Channon, 2008; Markman & Tetlock, 2000).

Experiment 1: Visible voting

In this experiment, participants examined voting scenarios where some voters' votes were visible to others and rated how responsible each contributing voter was for the outcome.

Methods

All experiments were pre-registered on the Open Science Framework and programmed with `jsPsych` (de Leeuw et al., 2023).[1]

[1]All data and materials can be found at: https://github.com/cicl-stanford/responsibility_influence_cogsci.

Participants Fifty participants (age: median = 40, SD = 11; gender: 27 females, 21 males, 1 non-binary and 1 undisclosed; race: 40 White, 6 Black, 3 Asian, 1 Multiracial) were recruited on Prolific. All participants were fluent English speakers based in the US, and were compensated at a rate of $12/hour.

Procedure Participants first clicked through instructions describing the study set-up, in which five club members voted on proposals and a majority was needed for a proposal to pass. Participants were familiarized with different types of scenarios: (1) trials in which all individuals voted independently; (2) trials in which one person's vote was visible to one or two others, and that person knew their vote would be seen; and (3) trials in which one person's vote was visible, but that person was unaware of their visibility. On each trial, observers (dependent causes) were illustrated as having open eyes (see Figure 1B). Agents who knew their votes would be seen (root causes with knowledge) were illustrated with thought bubbles. Arrows indicated the direction of observation by pointing from dependent voters to root voters. Each person's vote was indicated by a check mark or cross mark next to their name. After reviewing instructions, participants answered a set of comprehension questions. If they did not answer all the questions correctly, they would be redirected to the beginning of the instructions.

Participants then saw 23 trials in randomized order. On each trial, they rated how responsible each individual who voted with the majority was for the outcome, on a continuous slider with endpoints labeled "not at all" (0) and "very much" (10). For example, in Figure 1B, participants were asked to judge

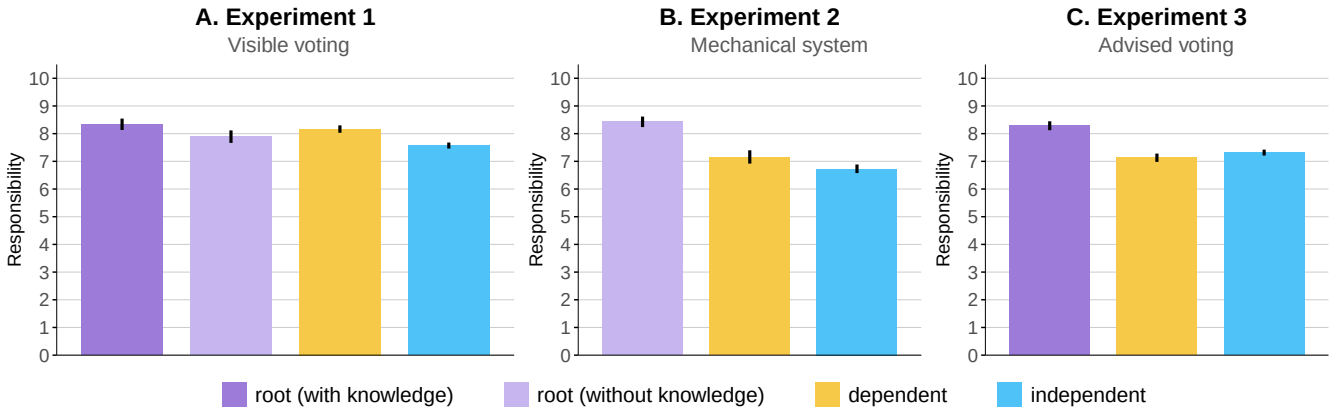


Figure 2: **Responsibility judgments by role**. Error bars are bootstrapped 95% confidence intervals. Root gears in Experiment 2 are referred to as root causes without knowledge because they are physical objects. Mentors in Experiment 3 are referred to as root causes with knowledge because they directly advise mentees.

responsibility for Sage, Hunter, Quinn, and Chandler (but not Bryce). After completing all trials, participants optionally shared demographic information and briefly described what factors influenced their responses. The average time to complete the experiment was 13 minutes (SD = 5.6).

Design We designed three different influence structures: either all voters were independent, there was one root voter and one dependent voter, or there was one root voter and two dependent voters (see Figure 1A for examples). This minimal design allowed us to isolate differences between roles and between root voters with different numbers of dependents. For the latter two structures, we additionally manipulated whether the root voter was aware that their vote was visible to others. We then created 23 trials that enumerate all possible voting patterns for these structures that lead to a positive outcome (i.e. a majority vote of 3, 4, or 5). Note that this means that there were some instances in which a dependent cause voted differently from the root cause, but across all scenarios they voted the same way as the root cause 73% of the time.

Results

We first evaluated how participants' responsibility judgments differed by the type of voter (root voter who knew their vote would be visible, root voter who was unaware their vote would be visible, dependent voter who observed a root voter, and independent voter). We fit a Bayesian mixed-effects model to predict responsibility ratings using role as a fixed effect. The model also controlled for majority size (i.e. number of voters in favor), and additionally included a random intercept for participant. Overall, we found credible effects of both the number of majority votes and a person's level of influence on how much responsibility they were assigned (Table 1). Responsibility decreased with the number of votes. That is, voters who were part of a larger majority were overall held less responsible than those who were part of a slight majority. To examine how responsibility for different types of voters compared with each other, we estimated marginal means and pairwise contrasts. Participants held root voters with knowledge the most responsible (M = 8.34, 95% CI

Table 1: **Model results**. Estimated coefficients for model predictors (size of majority; outcome in Experiment 3 only; and role) and pairwise differences between roles using marginal means, with 95% credible intervals. The labels root$^+$ and root$^-$ refer to root causes with and without knowledge, respectively. The size of majority refers to the number of contributors aligned with the outcome (3, 4, or 5). Credible effects are indicated with *.

Predictor	Estimate [95% CrI]
Experiment 1	
Size of majority	−0.07 [−0.13, −0.02] *
Role (root$^+$ vs. root$^-$)	0.44 [0.25, 0.62] *
Role (root$^+$ vs. dependent)	0.16 [0.01, 0.32] *
Role (root$^+$ vs. independent)	0.75 [0.60, 0.89] *
Role (dependent vs. root$^-$)	0.27 [0.11, 0.43] *
Role (dependent vs. independent)	0.59 [0.48, 0.69] *
Role (root$^-$ vs. independent)	0.31 [0.17, 0.46] *
Experiment 2	
Size of majority	−0.37 [−0.47, −0.28] *
Role (root$^-$ vs. dependent)	1.22 [0.98, 1.45] *
Role (root$^-$ vs. independent)	1.59 [1.40, 1.79] *
Role (dependent vs. independent)	0.38 [0.18, 0.57] *
Experiment 3	
Outcome	−0.06 [−0.15, 0.03]
Size of majority	−0.07 [−0.13, 0]
Role (root$^+$ vs. dependent)	1.15 [1.01, 1.29] *
Role (root$^+$ vs. independent)	0.96 [0.84, 1.08] *
Role (independent vs. dependent)	0.19 [0.09, 0.30] *

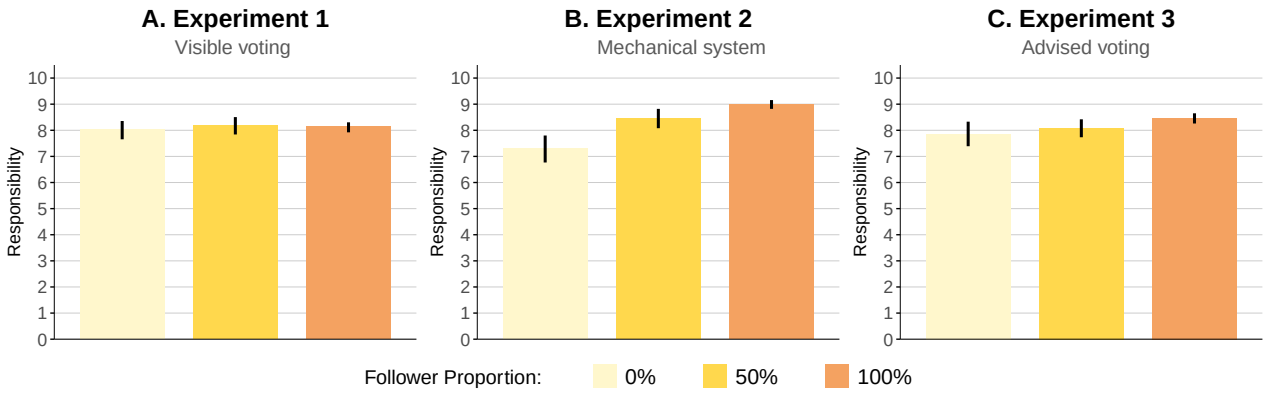


Figure 3: **Responsibility for root causes by follower proportion.** Error bars are bootstrapped 95% confidence intervals. In Experiments 1 and 3, followers refer to dependent voters who voted the same as the root voter. In Experiment 2, followers refer to dependent gears that spun with the root gear.

[8.13, 8.54]), followed by dependent voters (M = 8.17, 95% CI [8.03, 8.30]), then root voters without knowledge (M = 7.90, 95% CI [7.69, 8.12]), and finally independent voters (M = 7.57, 95% CI [7.46, 7.68]; Figure 2A). All differences in means were credible (Table 1).

We next tested how responsibility for root voters might vary with the number of dependent voters, particularly those that acted the same way. We fit a Bayesian mixed-effects regression to the subset of participants' ratings for root voters using a fixed effect of the proportion of dependent voters who followed the root's vote (either 0%, 50%, or 100%), as well as random intercepts for participants. We found no credible relationship between responsibility for root voters and the proportion of followers (β = 0.09, 95% CrI [−0.09, 0.27]). Participants assigned root voters similar amounts of responsibility regardless of whether none of their dependent voters voted the same way (M = 8.02, 95% CI [7.67, 8.37]), half of them followed (M = 8.18, 95% CI [7.82, 8.50]), or all of them followed (M = 8.13, 95% CI [7.95, 8.33]; Figure 3A).

Discussion

The results of Experiment 1 show that responsibility for group outcomes is affected by how individuals within the group influence each other, even when their contributions are the same. Voters who could be observed (root causes) were held the most responsible, especially if they knew about their own visibility. This finding was aligned with our initial hypotheses. However, voters who could observe others (dependent causes) were held next most responsible, and voters who acted independently were held the least responsible. This finding was surprising because we had expected that dependent voters would be perceived to have the least agency (Gray & Wegner, 2009) and smallest probabilistic impact on the outcome (Spellman, 1997), and thus be judged the least responsible.

A possible explanation is that dependent voters were not actually perceived to have less agency. In fact, some participants viewed dependent voters as having *greater* control over the outcome because they had access to more information (e.g. "When someone was able to see another vote, they had more power to swing the vote in favor or against."). In contrast, root voters without knowledge may have been held less responsible than dependent voters because, despite their influence, they did not have the additional information about others' votes. This explanation may also account for the null effect of follower proportion on responsibility for root causes. If dependent voters were simply making more informed decisions, then follower proportion would not necessarily bear on a root voter's causal efficacy. Based on these considerations, we designed Experiment 2 to eliminate confounds in how much information each contributor had by turning to a purely physical setting.

Experiment 2: Mechanical system

In Experiment 1, the relationship between root voters and dependent voters may have been interpreted as a difference in information access rather than direct causal influence. To remove mental states such as knowledge as potential confounds, we designed Experiment 2 to feature a fully mechanical system of gears.

Methods

Participants Fifty participants (age: median = 42, SD = 14; gender: 22 females, 28 males; race: 40 White, 6 Asian, 2 Black, 1 Hispanic, 1 Multiracial) were recruited on Prolific. All participants were fluent English speakers based in the US, and were compensated at a rate of $12/hour.

Procedure The procedure was similar to that of Experiment 1, except that we now framed the scenarios as systems of gears working together inside toy boxes (see Figure 1C). Each toy box consisted of five gears and at least three gears needed to spin for the toy box to work. Some gears (root gears) could cause others (dependent gears) to spin, although not always. Arrows indicated the direction of influence by pointing

from root gears to dependent gears. On each trial, participants were asked to rate how responsible each spinning gear was for the outcome on a continuous slider with endpoints labeled "not at all" (0) and "very much" (10). The average time to complete the experiment was 6.9 minutes (SD = 2.6).

Design The design was similar to that of Experiment 1, except that we no longer manipulated knowledge for the root cause. We continued to show only positive outcomes and the same three influence structures as before: all independent gears, one root gear with one dependent gear, or one root gear with two dependent gears. We created 14 trials that enumerate all possible gear-turning patterns that lead to a positive outcome. Across all trials, 65% of dependent gears spun when the root gear spun.

Results

As shown in Figure 2B, we first evaluated how participants' responsibility judgments differed by the role within the group (root gear without knowledge, dependent gear, or independent gear). We fit a Bayesian mixed-effects model to predict responsibility ratings using both role and size of majority (i.e. number of gears that spun) as fixed effects, along with a random intercept for participant. Overall, we found credible effects of both the size of the majority and a gear's level of influence (Table 1). Responsibility decreased as the total number of spinning gears increased. Participants held root gears the most responsible (M = 8.43, 95% CI [8.24, 8.61]), followed by dependent gears (M = 7.15, 95% CI [6.91, 7.39]), and independent gears (M = 6.73, 95% CI [6.57, 6.88]; Figure 2B). All pairwise differences were credible (Table 1).

We then tested whether responsibility for root gears varied with the number of dependent gears, particularly those that also spun. Participants held the root gear most responsible when 100% of dependent gears also spun (M = 9.00, 95% CI [8.81, 9.17]), followed by when 50% of dependent gears also spun (M = 8.45, 95% CI [8.07, 8.81]), and finally when none of the dependent gears spun (M = 7.30, 95% CI [6.84, 7.75]) (Figure 3B). We then fit a Bayesian mixed-effects regression to predict responsibility for root gears as a function of the proportion of spinning dependent gears (0%, 50%, or 100%). We found a credible positive relationship between responsibility for root gears and the proportion of dependent spinning gears (β = 1.66, 95% CrI [1.39, 1.92]).

Discussion

The results of Experiment 2 were similar to those in Experiment 1. Root gears were held the most responsible, and independent gears were held the least responsible. We additionally found that root gears were assigned more responsibility the more dependent gears spun with them. Interestingly, we see the same ordering of relative responsibility in dependent, independent, and root gears despite the setting here being mechanistic, which rules out explanations that dependent gears had more agency or knowledge over the outcome. One possibility here is that participants may have had strong prior intuitions about causal relationships between physical objects such as gears. They may have found trials in which the root gear spun, but the dependent gear (presumably linked in some way) did not spin, to be confusing. Relatedly, they may have perceived dependent gears as a unit with root gears so that their motion seemed tightly coupled. This could have led them to distribute responsibility across the unit, which would shift judgments for dependent gears and root gears to be in line with each other (e.g. "A lot of cases were pretty straightforward. The hardest ones tended to be the ones where gears failed to turn other ones. It felt like this caused [root gears] to not contribute a lot to the final outcome.").

We saw that in Experiment 1, dependent voters may have been perceived to be merely more informed agents, while in Experiment 2, dependent gears may have been seen as inextricably linked to root gears. Therefore, we designed Experiment 3 to test a new framing that preserves the agency of the contributors but makes social influence more explicit.

Experiment 3: Advised voting

In Experiment 3, we returned to the voting paradigm from Experiment 1, but framed influence explicitly as mentors giving advice to mentees before voting. With this setup, mentors act as root causes with knowledge.

Methods

Participants Fifty participants (age: median = 43, SD = 11; gender: 24 females, 26 males; race: 39 White, 4 Black, 3 Asian, 3 Multiracial, 1 Middle Eastern) were recruited on Prolific. All participants were fluent English speakers based in the US, and were compensated at a rate of $12/hour.

Procedure The procedure was similar to that of Experiments 1 and 2, except that we now framed influence as a mentor-mentee relationship. Some members in each group were mentees (dependent causes), who received advice from a mentor (root cause) prior to everyone privately casting their votes (see Figure 1D). Arrows pointed from root causes to dependent causes. In the main task, participants were shown 20 trials in randomized order. For each trial, they rated how responsible each member who voted with the majority was for the proposal passing or failing to pass. The average completion time was 10.2 minutes (SD = 4).

Design Unlike the previous two experiments, here we included trials with both positive and negative outcomes. We created 20 trials (12 positive, 8 negative) that collectively covered a wide variety of situations differing in the proportion of dependent voters who followed the root voter's advice, the number of dependent voters, and how votes were distributed across group members. Across all trials, dependent voters followed the root voters' advice 70% of the time.

Results

Figure 2C shows mean responsibility judgments by voter type. Using the same Bayesian mixed-effects models as in Exper-

iments 1 and 2, we examined responsibility judgments as a function of role (root voter with knowledge, i.e. mentor; dependent voter, i.e. mentee; and independent voter), controlling for outcome (positive or negative) and majority size, and including random intercepts for participants. We did not find credible effects of outcome or majority size on responsibility judgments (Table 1).

We found a credible effect of a person's level of influence on how much responsibility they were assigned. Participants held root voters with knowledge most responsible (M = 8.29, 95% CI [8.13, 8.45]), followed by independent voters (M = 7.32, 95% CI [7.20, 7.43]), and lastly dependent voters (M = 7.14, 95% CI [6.97, 7.28]; Figure 2C). All pairwise differences were credible (Table 1).

Responsibility for root voters varied with the number of dependent voters who acted the same way as them (Figure 3C). We found a credible positive relationship between root voter responsibility and the proportion of dependent voters that followed (β = 0.64, 95% CrI [0.38, 0.90]). Participants assigned the most responsibility to root voters when 100% of dependents followed them (M = 8.47, 95% CI [8.27, 8.66]), and next when half of them followed (M = 8.09, 95% CI [7.75, 8.42]), and finally the least responsibility when none of them did (M = 7.86, 95% CI [7.37, 8.34]).

Discussion

In this experiment, we studied responsibility among agents where we explicitly stated the presence and direction of influence as a mentor advising a mentee. Consistent with prior results, we found that mentors (root causes) were attributed the most responsibility and that they were held more responsible when more of their mentees followed their advice.

In contrast to Experiments 1 and 2, here we found that dependent causes were held less responsible than independent causes. This reversal highlights the nuanced factors that affect responsibility judgments, including inferences about a person's mental states such as intentionality and foreseeability (e.g. Cushman, 2008; Lagnado & Channon, 2008; Markman & Tetlock, 2000), as well as perceptions of agency and personal control (e.g. Alicke, 2000; Gray & Wegner, 2009). In Experiment 1, where root voters were merely observed, their intentions to influence were not clear (especially when they were unaware) and there were no particular relationships to suggest that dependent voters might have less agency. In Experiment 3, however, mentors actively and intentionally advised mentees, forming a stronger impression of root causes causally shaping the contributions of dependent causes.

General Discussion

People's actions are rarely solitary endeavors, but instead informed and interpreted by those around us. In this paper, we investigated how people attribute responsibility for group outcomes when some individuals in the group can influence the actions of others. Across three experiments spanning both social and physical settings, we found that responsibility judgments are affected by whether a person can influence others (a root cause), is influenced by others (a dependent cause), or acts independently (an independent cause), even when all contributions to the outcome are identical.

First, we found that root causes were consistently held more responsible than both dependent and independent causes. Their responsibility tended to increase with the proportion of dependent causes that followed. This aligns with prior findings that people can be held responsible for setting bad examples (Kardos et al., 2024) or by being associated with wrongdoers (Lickel & Onuki, 2015). Second, an intriguing finding was how responsibility for dependent causes shifted relative to independent causes across different contexts. Participants held dependent causes more responsible when influence was more passive (Experiment 1) or mechanistic (Experiment 2), but reversed this pattern when influence was highlighted as a distinct social relationship (Experiment 3). One explanation is that this reflects a distinction between actively shaping another's behavior versus merely permitting or enabling the possibility of an outcome. Different causal interpretations and expressions can produce different causal judgments (Beller & Gerstenberg, 2026; Wolff, 2007).

In an exploratory follow-up analysis, we re-ran all mixed-effects models with random slopes for all within-participant predictors. We found that, in Experiment 1, only the differences between root causes (both with and without knowledge) and independent causes remained credible; differences between dependent causes and any other causes were no longer credible. In Experiments 2 and 3, root causes were still held credibly more responsible than dependent and independent causes, but differences between the latter two were no longer credible. Future work should clarify these effects.

Overall, our findings also raise interesting questions about how different forms of influence shape responsibility judgments. In our experiments, dependent causes generally followed root causes, but oftentimes people are influenced to act in the *opposite* direction (e.g. Brehm, 1966). For example, while a 2024 poll found that 18% of people were more likely to vote for a political candidate endorsed by Taylor Swift, 17% reported they would be *less* likely to do so (Fung, 2024). How do people judge responsibility in settings with mixed or even conflicting social pressures? A related case is when a root has the capacity to influence others but chooses not to use it, either deliberately or negligently. Prior work suggests that acts of omission readily elicit blame, especially when the agent had a duty or normative prior to act (DeScioli et al., 2012; Haidt & Baron, 1996; Henne et al., 2017; Khemlani et al., 2021; Sarin & Cushman, 2024; Spranca et al., 1991). Even awareness of one's potential influence may create a moral obligation to exercise it (Hart & Honoré, 1985; Singer, 1972). Future work could examine responsibility attribution in these cases.

Together, the current work contributes towards a better understanding of responsibility attribution that accounts not only for the direct impacts of our individual actions, but also for how our actions influence others' propensity to act.

Acknowledgments

We thank members of the Causality and Cognition Lab for helpful discussion and feedback. TG was supported by grants from the Stanford Institute for Human-Centered Artificial Intelligence (HAI) and from the Cooperative AI Foundation. SAW was supported by a Stanford Interdisciplinary Graduate Fellowship.

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9), 1082–1095.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In *Groups, leadership and men: Research in human relations* (pp. 177–190). Carnegie Press.

Beller, A., & Gerstenberg, T. (2026). Causation, meaning, and communication. *Psychological Review*, *133*(2), 339–381.

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

Brewer, M. B. (1977). An information-processing approach to attribution of responsibility. *Journal of Experimental Social Psychology*, *13*(1), 58–69.

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, *22*, 93–115.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380.

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.

DeScioli, P., Asao, K., & Kurzban, R. (2012). Omissions and byproducts across moral domains. *PLOS ONE*, *7*(10), e46963.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636.

Duffy, K. A., & Chartrand, T. L. (2015). Mimicry: Causes and consequences. *Current Opinion in Behavioral Sciences*, *3*, 112–116.

Engelmann, N., & Waldmann, M. R. (2022). How causal structure, causal strength, and foreseeability affect moral judgments. *Cognition*, *226*, 105167.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Fung, K. (2024). Can Taylor Swift sway voting in the 2024 election? *Newsweek*.

Gantman, A. P., Sternisko, A., Gollwitzer, P. M., Oettingen, G., & Van Bavel, J. J. (2020). Allocating moral responsibility to multiple agents. *Journal of Experimental Social Psychology*, *91*, 104027.

Gerstenberg, T., & Lagnado, D. A. (2010). Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition*, *115*(1), 166–171.

Gerstenberg, T., Lagnado, D. A., & Zultan, R. (2023). Making a positive difference: Criticality in groups. *Cognition*, *238*, 105499.

Gerstenberg, T., Ullman, T. D., Nagel, J., Kleiman-Weiner, M., Lagnado, D. A., & Tenenbaum, J. B. (2018). Lucky or clever? From expectations to responsibility judgments. *Cognition*, *177*, 122–141.

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520.

Haidt, J., & Baron, J. (1996). Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology*, *26*(2), 201–218.

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part II: Explanations. *British Journal for the Philosophy of Science*, *56*(4), 889–911.

Hart, H. L. A., & Honoré, T. (1985). *Causation in the law* (2nd ed.). Oxford University Press UK.

Henne, P., Pinillos, Á., & De Brigard, F. (2017). Cause by omission and norm: Not watering plants. *Australasian Journal of Philosophy*, *95*(2), 270–283.

Hilton, D. J., McClure, J. L., & Sutton, R. M. (2010). Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes? *European Journal of Social Psychology*, *40*(3), 383–400.

Kardos, P., Leidner, B., & Lickel, B. (2024). The responsibility of setting a bad example: Models are blamed for their imitators' behaviour. *European Journal of Social Psychology*, *54*(7), 1500–1514.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128.

Khemlani, S., Bello, P., Briggs, G., Harner, H., & Wasylyshyn, C. (2021). Much ado about nothing: The mental representation of omissive relations. *Frontiers in Psychology*, *11*.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754–770.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, *37*(6), 1036–1073.

Langenhoff, A. F., Wiegmann, A., Halpern, J. Y., Tenenbaum, J. B., & Gerstenberg, T. (2021). Predicting responsibility judgments from dispositional inferences and causal attributions. *Cognitive Psychology*, *129*, 101412.

Lickel, B., & Onuki, M. (2015). Generalization processes in collective responsibility and intergroup conflict. In *Social perception from individuals to groups* (pp. 197–212). Psychology Press.

Lickel, B., Schmader, T., & Hamilton, D. L. (2003). A case of collective responsibility: Who else was to blame for the Columbine High School shootings? *Personality and Social Psychology Bulletin*, *29*(2), 194–204.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *The Journal of Psychology*, *25*(1), 147–186.

Markman, K. D., & Tetlock, P. E. (2000). 'I couldn't have known': Accountability, foreseeability and counterfactual denials of responsibility. *British Journal of Social Psychology*, *39*(3), 313–325.

McClure, J. L. (1998). Discounting causes of behavior: Are two reasons better than one? *Journal of Personality and Social Psychology*, *74*(1), 7–20.

Sanders, J., Lee Hamilton, V., Denisovsky, G., Kato, N., Kawai, M., Kozyreva, P., Kubo, T., Matskovsky, M., Nishimura, H., & Tokoro, K. (1996). Distributing responsibility for wrongdoing inside corporate hierarchies: Public judgments in three societies. *Law & Social Inquiry*, *21*(4), 815–855.

Sarin, A., & Cushman, F. (2024). One thought too few: An adaptive rationale for punishing negligence. *Psychological Review*, *131*(3), 812–824.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer Science & Business Media.

Shultz, T. R., Jaggi, C., & Schleifer, M. (1987). Assigning vicarious responsibility. *European Journal of Social Psychology*, *17*(3), 377–380.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy and Public Affairs*, *1*(3), 229–243.

Spellman, B. A. (1997). Crediting Causality. *Journal of Experimental Psychology: General*, *126*(4), 323–348.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*(1), 76–105.

Sullivan, B. (2023). A Taylor Swift Instagram post helped drive a surge in voter registration. *NPR*.

van de Ven, N., Archer, A. T. M., & Engelen, B. (2019). More important and surprising actions of a moral exemplar trigger stronger admiration and inspiration. *The Journal of Social Psychology*, *159*(4), 383–397.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psycholog:. General*, *136*, 82–111.

Wu, S. A., & Gerstenberg, T. (2024). If not me, then who? Responsibility and replacement. *Cognition*, *242*, 105646.

Wu, S. A., Sridhar, S., & Gerstenberg, T. (2023). A computational model of responsibility judgments from counterfactual simulations and intention inferences. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*.

Xiang, Y., Landy, J., Cushman, F. A., Vélez, N., & Gershman, S. J. (2023). Actual and counterfactual effort contribute to responsibility attributions in collaborative tasks. *Cognition*, *241*, 105609.