Responsibility judgments in voting scenarios

Tobias Gerstenberg[1] (tger@mit.edu) **Joseph Y. Halpern[2] (halpern@cs.cornell.edu)**
Joshua B. Tenenbaum[1] (jbt@mit.edu)
[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139
[2]Computer Science Department, Cornell University, Ithaca, NY 14853

Abstract

How do people assign responsibility for the outcome of an election? In previous work, we have shown that responsibility judgments in achievement contexts are affected by the probability that a person's contribution is necessary, and by how close it was to being pivotal (Lagnado, Gerstenberg, & Zultan, 2013). Here we focus on responsibility judgments in voting scenarios. We varied the number of people in different voting committees, their political affiliations, the number of votes required for a policy to pass, which party supports the policy, and the pattern of votes (creating 170 different situations). As expected, we found that participants' responsibility judgments increased the closer the voter was to being pivotal. Further, judgments increased the more unexpected a vote was. Voters were assigned more responsibility when they voted against the majority in the committee, and when they voted against their party affiliation.

Keywords: responsibility, causality, counterfactuals, pivotality, criticality, normality, voting.

Introduction

How do people assign responsibility to individuals for a group outcome? Intuitively, responsibility is closely connected to difference-making. In order to be held responsible for an outcome, one's action must in some way have made a difference to the outcome. However, there are many situations in everyday life in which an individual action doesn't make a difference. For example, when we vote, the chance that our individual vote will be pivotal is marginal. Indeed, the fact that so many people actually vote despite the small chance that their individual vote will make a difference has puzzled economists attempting to provide a rational explanation of voting behavior. Taking up the individual costs of commuting to the polling station and standing in line (sometimes for hours) doesn't seem to be justified by the minuscule chances of casting the pivotal vote. Goldman (1999) has argued that taking into account responsibility can explain why people vote. If we can get partial credit for positive outcomes (and partial blame for negative ones), then going voting maximizes one's chances to receive (partial) credit and minimizes the chances to receive (partial) blame. People's concern with how their actions are evaluated by others thus provides a reason to vote (see also Scanlon, 2009).

We sympathize with Goldman's (1999) account. In this paper, our aim is to better understand exactly how people assign responsibility in voting scenarios. Goldman (1999) did not provide a model of partial credit or blame. We develop such a model and test how well it explains participants' responsibility judgments.

Pivotality & Responsibility How can we capture whether a person's action made a difference to the outcome? Halpern and Pearl (2005) provide a definition of causality in Pearl's (2000) structural-model framework. The definition involves a counterfactual contrast: we compare the actual outcome with what the outcome would have been if the person's action had been different. However, such a naive use of counterfactuals does not suffice in general as a model of responsibility. Consider a very simple voting scenario with two members in a committee, Jack and Bill, who vote on whether or not a certain policy should be passed. In order for the policy to pass, at least one of the committee members has to vote in its favor. If both of the members vote against the policy, it won't be passed. In fact, both Jack and Bill ended up voting for the policy. Here, we have a simple situation in which the outcome was overdetermined and neither of the individual actions made a difference to the outcome. Even if Jack had voted against the policy, it would still have been passed due to Bill's vote. However, intuitively Jack and Bill are still (at least partially) responsible for the policy having been passed even though each person's action made no difference in the actual situation.

Halpern and Pearl (2005) deal with this problem by employing a more relaxed test for counterfactual dependence. Their definition makes Jack and Bill causes in the case of overdetermination, however, it does not distinguish the degree of responsibility of Jack and Bill if the vote is 2–0 or if the vote is 10–0. In both cases, Jack and Bill are causes. Chockler and Halpern (2004) refine the Halpern-Pearl notion of causality by defining a notion of *degree of responsibility*. A person's responsibility decreases the "further away" his action was from having made a difference to the outcome. The greater the number of changes required to move from the actual situation to a situation in which the person's action was pivotal, the less responsible the person is predicted to be seen.

We call this notion the *pivotality* of a person's action A in a given situation S for a particular outcome E. Formally, pivotality is defined as

$$\mathit{Pivotality}(A,S,E) = \frac{1}{C+1}, \tag{1}$$

where C is the minimal number of changes that are required to make A pivotal for E in S.[1] In the voting scenarios that we consider, C simply represents the number of other voters who would have needed to vote differently in order for the person under consideration to become pivotal. Thus, Jack's

[1] S captures the causal structure of the situation, which is often represented in terms of structural equations. For our voting scenarios, S captures the threshold of votes required in order for a policy to be passed.

pivotality in the example above is $\frac{1}{2}$ ($\frac{1}{1+1}$), since Bill's vote needs to be changed to make Jack pivotal.

Criticality & Responsibility In previous work, we tested the pivotality model in achievement contexts in which participants assigned responsibility to individual members for the outcome of their team (Gerstenberg & Lagnado, 2010, 2012; Lagnado et al., 2013; Zultan, Gerstenberg, & Lagnado, 2012). As predicted by the model, these experiments showed that people's responsibility judgments are sensitive to how close a person was to being pivotal. However, the experiments also revealed a pattern of judgments that cannot be explained merely in terms of pivotality. We contrasted situations in which the task was conjunctive (all of the members need to do well in order for the team to succeed) versus disjunctive (at least one of the members needs to do well). Consider a situation with two team members who both failed in their task. If the task was disjunctive, then each of the team members was pivotal; the team would have succeeded had either of them passed their task. If the task was conjunctive, then the pivotality of each failed member was reduced to $\frac{1}{2}$. However, participants' responsibility judgments showed the opposite pattern: they assigned more responsibility to a team member when both failed in the conjunctive task than in the disjunctive task.

In order to explain this pattern of results, we postulated that people care not only about how close a person's action was to having been pivotal *ex post*, but also about how critical the person was *a priori*. We define the criticality of a person P in situation S as

$$Criticality(P,S) = 1 - \frac{p(E|\neg A)}{p(E|A)}, \tag{2}$$

where $p(E|A)$ is the probability of a positive team outcome if P's action succeeded, and $p(E|\neg A)$ is the probability that the team will succeed if P's action failed. In a conjunctive task, each person's contribution is critical. If any of the team members fails in their task, the probability of the team succeeding is 0 (i.e., $p(E|\neg A) = 0$). In contrast, in disjunctive situations, each team member's criticality is reduced. If we assume that each player has a $p = 0.5$ chance of succeeding, then a team member's criticality in a team of two is $1 - \frac{0.5}{1} = 0.5$. Lagnado et al. (2013) experimentally varied criticality through creating different team tasks (disjunctive, conjunctive, and mixed) and pivotality through the performances of each player in the team (i.e., who succeeded and who failed). The results showed that both aspects had a significant influence on participants' judgments.

Normality & Responsibility So far we have identified pivotality and criticality as factors that influence people's responsibility judgments. It has also been shown that people's causal and responsibility judgments are influenced by normative considerations (Gerstenberg, Ullman, Kleiman-Weiner, Lagnado, & Tenenbaum, 2014; Hitchcock & Knobe, 2009; Kominsky, Phillips, Knobe, Gerstenberg, & Lagnado, 2015). Generally, people who acted against a norm are judged to be more causal for an outcome than people whose action was in line with a norm (e.g., Knobe & Fraser, 2008). Recently, formal models of actual causation have been extended to include normality considerations to better account for people's causal judgments (Halpern & Hitchcock, 2015).

Consider a situation in which Jack is a Republican and Bill is a Democrat, and both are in a committee voting on a policy that is supported by the Republican party. In this situation, Jack would be expected to vote for the policy and Bill to vote against it. Let us call this aspect of normality *dispositional normality*; it expresses our expectation of how a person will vote based on their political affiliation. We define the dispositional normality of an action A, taken by a committee member with a certain party affiliation M_p in a situation in which the policy is supported by party P_p as

$$Normality_D(A, M_p, P_p) = \begin{cases} 1 & \text{if } M_p = P_p \\ 0 & \text{if } M_p \neq P_p. \end{cases} \tag{3}$$

That is, a person's action is normal if they voted in line with their party affiliation and abnormal otherwise.

In voting scenarios, there is also another sense in which a person's action can be more or less normal. Consider a committee of five in which everyone except for Jack voted in favor of the policy. Here, Jack's action is less normal than it would be in a situation in which everyone else also voted against the policy. We call this aspect *situational normality*; it captures the sense in which a person's action in a situation S was how we expected it to be, given how others behaved in S. We define the situational normality of a person's action A in situation S when the outcome was E as

$$Normality_S(A, S, E) = \frac{\sum_{i=1}^{N} \mathbb{1}(A = A_i)}{N}, \tag{4}$$

where N equals the number of committee members, A_i is the vote of committee member i, and $\mathbb{1}(A = A_i) = 1$ if $A = A_i$ and 0 otherwise. For example, in a situation in which Jack voted for a policy, but the remaining four committee members voted against it, the situational normality of Jack's vote was $\frac{1}{5}$. Note that whereas the notions of criticality and dispositional normality are determined *before* the outcome is known, pivotality and situational normality take into account what actually happened in the particular situation.

Predictions

Based on previous research, we predict that participants' responsibility judgments to voters in a committee increase the closer their vote was to having been pivotal. Voters should also be judged more responsible to the extent that their vote was perceived to be critical. Finally, we predict that both dispositional and situational normality influence participants' judgments. A voter should be judged more responsible to the extent that their vote was perceived to have been abnormal.

While the notion of pivotality as defined above is orthogonal to criticality and normality, the latter two notions are

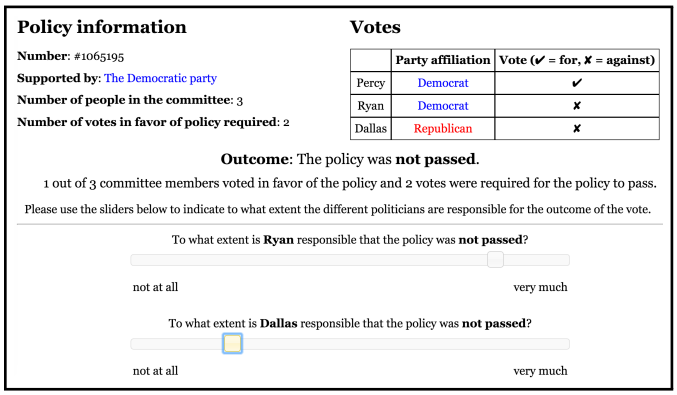

Figure 1: Experiment screenshot.

related. Consider again, the example of Jack, the Republican, and Bill, the Democrat, voting on a policy supported by the Republican party which needs at least one vote in order to be passed. Jack is more critical than Bill assuming that, a priori, a Republican has a higher probability of voting for the policy than a Democrat does. Let's assume that the probability that Jack and Bill will vote in favor of the policy is $p_J = 0.75$ and $p_B = 0.25$, respectively. Plugging these values into Equation 2, we get that Jack's criticality is $1 - \frac{0.25}{1} = 0.75$, and Bill's criticality is $1 - \frac{0.75}{1} = 0.25$.

Let's consider a situation in which both Jack and Bill voted against the policy. In this situation, both criticality and dispositional normality considerations predict that Jack will be seen as more responsible than Bill. Jack was more critical and his vote was also more abnormal. Now consider a situation in which both Jack and Bill voted for the policy. Since a person's criticality is determined a priori, it is not affected by the outcome. Jack was still more critical than Bill. However, in this case, Bill's vote was more abnormal than Jack's. When the outcome is negative both criticality and dispositional normality pull in the same direction. However, when the outcome is positive, criticality and dispositional normality pull in opposite directions. Consequently, we predict that a voter's party affiliation will have a stronger effect on people's responsibility judgments when a policy wasn't passed than when a policy was passed.

Experiment

In the experiment, participants' task was to assign responsibility to committee members for the outcome of a vote on a policy. Figure 1 shows an example situation. Policy #1065195, which was supported by the Democratic party was up for vote. There were three people on the committee: two Democrats, Percy and Ryan, and one Republican, Dallas. At least two votes in favor of the policy were required in order for the policy to be passed. As it turned out, only Percy voted in favor of the policy while Ryan and Dallas voted against the policy. The policy was not passed since two votes would have been required but only one committee member voted for the policy.

Methods

Participants 208 participants (M_{age} = 36.24, SD_{age} = 13.54, 86 female) were recruited via Amazon Mechanical Turk. Participation was restricted to workers based in the US with a prior approval rate greater than 95% (Mason & Suri, 2012).

Design Table 1 shows some of the patterns that were used in the experiment. We manipulated the size of the committee ($N = 3$ vs. $N = 5$), the political affiliations of the committee members (M_{p_i}), how each committee member voted (v_i), and the threshold for the policy to be passed (T). For example, #16 is the situation shown on the screenshot in Figure 1.

In principle, there would have been $2^3 \times 2^3 \times 3 + \times 2^5 \times 2^5 \times 5 = 5312$ different possible situations, taking into account the political affiliations, pattern of votes, and the different thresholds for committees of size 3 and 5. However, since the votes are being cast simultaneously, there are many situations that are symmetrical for our purposes. For example, if all of the committee members were Democrats, and two voted for the policy while one voted against it, we don't care about which out of the three it was that voted against the policy. Taking into account these symmetries already significantly reduces the number of situations to 340.

We further reduced the number of situations by removing all situations for which the pattern of votes was unusual. We defined a situation to be unusual when a majority of the committee voted against their political affiliation. For example, consider a situation in which the policy is supported by the Democrats but all committee member are Republicans. Here, we removed all the situations in which more than 2 of the Republicans voted in favor of the policy. Removing all unusual situations reduces the number of situations to 170 (30 situations for committees of size 3, and 140 situations for committees of size 5).

We split the 170 situations into 10 different conditions with 17 situations each. Each condition included 3 situations with $N_{committee} = 3$, and 14 situations with $N_{committee} = 5$.

Procedure

Participants were randomly assigned to one of 10 conditions. After receiving instructions, each participant made responsibility judgments for a set of 17 situations. Participants judged to what extent a particular committee member was responsible that the policy passed (or didn't pass; see Fig-

Table 1: Examples of patterns for the situations with $N = 3$ committee members. *Note*: M_{p_i} = political affiliation: 0 = opposite from the party which supports the policy, 1 = same party; v_i = vote: 0 = against, 1 = for, S = sum of votes in favor; T = threshold for policy to be passed; O = outcome: 0 = policy was not passed, 1 = policy was passed.

#	M_{p_1}	M_{p_2}	M_{p_3}	v_1	v_2	v_3	S	T	O
1	0	0	0	0	0	0	0	1	0
				...					
16	1	1	0	1	0	0	1	2	0
				...					
30	1	1	1	1	1	1	3	3	1

ure 1). Participants made their judgments on sliding scales whose endpoints were labeled with "not at all" (0 responsibility) and "very much" (100 responsibility).

Participants were asked only to assign responsibility to committee members whose vote was in line with the outcome. Depending on the situation, participants were either asked to make one or two judgments. When all committee members whose vote was in line with the outcome shared the same party affiliation, participants made only one judgment. When at least two of the committee members whose vote was in line with the outcome came from different political parties, then participants were asked to judge the responsibility for one of the Democrats and one of the Republicans. Out of the set of 170 situations, there were 90 situations in which participants were asked to make a single judgment, and 80 situations in which they made responsibility judgments for two committee members. Thus, we have a total of 250 data points.

For example, in situation #16 (depicted in Figure 1), because the two voters whose vote was in line with the outcome came from different political parties (Ryan and Dallas), participants were asked to judge the responsibility for each of them. Since Percy voted in favor of the policy, participants were not asked to judge to what extent he was responsible for the policy not being passed.

On average, it took participants 6.61 minutes ($SD = 7.03$) to complete the experiment.

Results

Figure 2 shows participants' mean responsibility judgments for a selection of cases. In Figure 2a, committee member 1 was judged very responsible for the policy being passed. In this situation, one vote was required for the policy to pass (Threshold = 1), member 1 voted in favor of the policy and the other two voted against it. The vote was pivotal, he voted against how the other two voters voted, but in line with his political affiliation. In Figures 2b) and c), all three committee members were from the same party that supported the policy and voted in favor of it. What changed between the situations was the number of votes required in order for the policy to pass. In Figure 2b) only one vote was required, while in Figure 2c) all three votes were required. Thus, the normality of the committee member's vote was constant between situations but the pivotality was different. In Figure 2b), the outcome was overdetermined, and thus the committee member's pivotality reduced (both of the other members would have needed to change their vote in order for the first committee member to have been pivotal).

Figures 2d) and e) show situations in which a policy failed to pass and members of different party affiliations voted against the policy. In both situations, the member who was from the party that supported the policy received more responsibility than the member who was from the opposite party. The difference is more marked in Figure 2e) than in Figure 2d). Finally, in Figure 2f), four out of five members voted in favor of a policy that needed four votes in order to be passed. In this situation, participants assigned slightly more

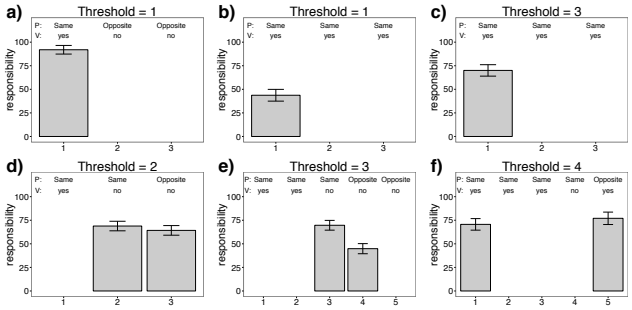

Figure 2: Mean responsibility judgments for a selection of cases. *Note*: P = political affiliation (same = from party which supports the policy, opposite = not from party which supports the policy), V = whether a committee member voted for or against the policy. Error bars denote $\pm 1SEM$.

responsibility to the opposite-party member than the same-party member.

Because of the large number of situations that participants saw in the experiment (170 different situations across 10 conditions), we cannot discuss each situation individually. We will now examine the data on a higher level of aggregation to see whether, and to what extent, participants' judgments were influenced by pivotality, normality, and criticality.

Pivotality Figure 3 shows participants' mean responsibility attributions as a function of pivotality. Let's consider situations in which the threshold T was 1 (i.e., the red line). In this case, there is only one way for the policy not to be passed: all members must have voted against the policy. In this situation, each of the members is pivotal; they could have changed the outcome had they voted differently.

In contrast, there are many different ways for the policy to be passed. Let's focus on situations in which there were $N = 5$ committee members. When the voter's pivotality was 1, it means that all other members voted against the policy. On the other extreme, when pivotality was 0.2, it means that

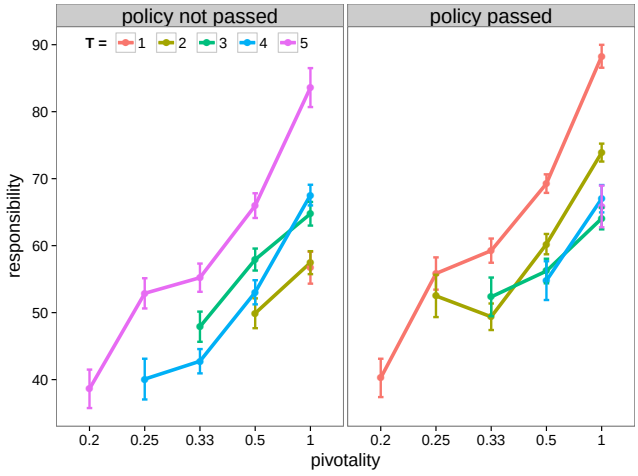

Figure 3: Mean responsibility judgments as a function of pivotality (x-axis) and the threshold (T) for the policy to pass (lines), separated for positive and negative outcomes (columns). Error bars denote $\pm 1SEM$.

all of the committee members voted in favor of the policy.

As predicted, responsibility reduces the further away a voter was from having been pivotal. This general trend holds irrespective of what the threshold of required votes was to pass the policy. Responsibility increases with pivotality for each line in Figure 3. However, there is another trend apparent in the data that cannot be explained merely in terms of pivotality considerations.

Normality: The Situation If pivotality was the only factor that influenced participants' judgments, then varying the threshold while keeping pivotality fixed shouldn't make any difference. However, if we compare participants' responsibility judgments in Figure 3 for different thresholds at the same level of pivotality, we see that the judgments are not identical.

For example, let's focus on situations in which the committee member was pivotal but the thresholds were different. When the policy didn't pass, participants attributed more responsibility to a member when the threshold was 5 than when it was lower. Conversely, when the policy was passed, a member was judged most responsible when the threshold was 1 compared to situations in which the threshold was higher.

We take this pattern of results to support the effect of situational normality considerations on participants' responsibility judgments. A member was judged most responsible when their vote was different from the votes of all the other committee members. When the policy was not passed, the member was judged more responsible when the threshold was 5 (which means that all others voted for the policy) than when the threshold was one (which implies that all members voted against the policy). Conversely, when the policy was passed, the member was judged most responsible when the threshold was 1 (which implies that all others voted against the policy) than when the threshold was 5 (meaning that all members voted for the policy).

So far, we have seen that both pivotality and the extent to which a person's vote was in line with other people's votes affect responsibility judgments.

Normality & Criticality: The Person We now examine the effect that manipulating party affiliation had on participants' judgments. We hypothesized that members whose vote was not in line with what would be expected from their party affiliation, will be judged more responsible for the outcome than members who voted as expected. Figure 4 shows participants' responsibility judgments as a function of pivotality and party affiliation. The effect of pivotality is consistent across members with different party affiliations.

To look at the effects of party affiliation more closely, we compared participants' responsibility judgments as a function of party affiliation in the subset of situations in which they were asked to make a judgment for one committee member of each party. In situations in which the policy was not passed, participants assigned significantly more responsibility to committee members from the party that generally supported the policy compared to opposite-party members ($t(207) = 2.58, p = .01, r = 0.13$). In situations in which the policy was passed, there was no significant difference in responsibility judgments to same versus opposite-party members ($t(207) = 0.73, p = .48, r = 0.04$).

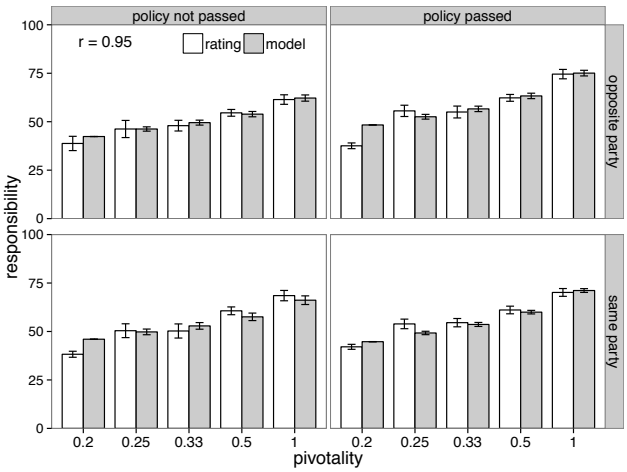

Figure 4: Mean responsibility judgments (white bars) and regression model predictions (gray bars, Model 3 in Table 2) as a function of pivotality (x-axis), outcome (columns), and party affiliation (rows). Error bars denote $\pm 1 SEM$.

As hypothesized above, the effect of party affiliation shows up in situations in which the policy wasn't passed, where both normality and criticality considerations predict that same-party members should receive more responsibility. In contrast, when a policy was passed, and dispositional normality and criticality pull in opposite directions, there was no significant effect of party affiliation on responsibility judgments.

Regression analysis Up until now, we have discussed the influence of pivotality, normality, and criticality on participants' responsibility judgments mostly qualitatively. We will now investigate to what extent the different factors we've identified, accurately capture participants' judgments by using them as predictors in different regression models (see Table 2). All models consider pivotality as a predictor. Model 1 additionally includes criticality as predictor. Model 2 includes dispositional and situational normality. In addition to criticality as well as both types of normality, Model 3 includes an outcome term that captures whether participants have a tendency to attribute more responsibility for positive or negative outcomes.

When only combined with pivotality, criticality is not a sig-

Table 2: Regression models.

	Model 1	**Model 2**	**Model 3**
Pivotality	24.11^{***}	15.95^{***}	15.14^{***}
Criticality	-1.16		6.66^{***}
Normality_D		-3.67^{***}	-3.66^{***}
Normality_S		-26.82^{***}	-27.03^{***}
Outcome			8.93^{***}
Constant	45.16^{***}	71.02^{***}	63.35^{***}
R^2	0.33	0.54	0.62
Res Std. Error	10.64	8.83	8.08
F Statistic	59.71^{***} (df = 2; 247)	95.40^{***} (df = 3; 246)	78.42^{***} (df = 5; 244)
Note:			*p<0.1; **p<0.05; ***p<0.01

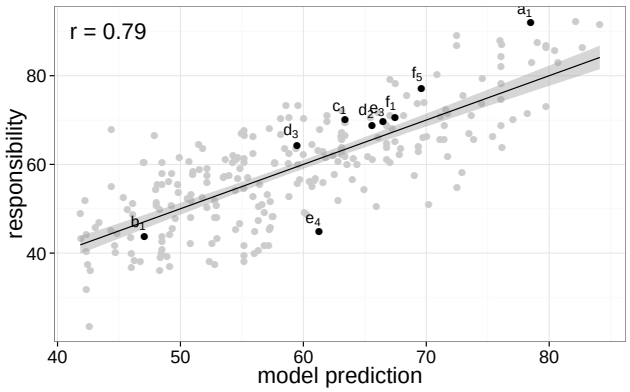

Figure 5: Correlation between model prediction (Model 3 in Table 2) and mean responsibility judgments for the 250 data points. *Note*: The labeled black dots refer to the cases shown in Figure 2. For example, e_3 refers to the judgment for committee member 3 in Figure 2e.

nificant predictor (Model 1), while both types of normality influence participants' judgments in the predicted direction. Participants assign more responsibility the more abnormal a person's action was (Model 2). Finally, the complete model (Model 3) shows that criticality does become a significant predictor of participants' judgments when it is combined with the normality predictors. Furthermore, it turns out that participants generally assigned more responsibility for committee members who voted in favor of a policy than for committee members who voted against a policy.

Figure 5 shows a scatter plot of the model predictions and participants' responsibility judgments for the full set of 170 situations (with 250 judgments).

Discussion

In this paper, we investigated how people assign responsibility to individuals for collective outcomes. We looked into a large number of voting scenarios that differed in terms of the size of the committee, the threshold required for a policy to pass, the political affiliations of the committee members, and the patterns of votes. The voting setup allowed us to manipulate pivotality, criticality, and normality in quantitative ways and see how people's judgments are affected by these different factors.

In line with previous work (Lagnado et al., 2013), the results showed that the closer a committee member's vote was to having been pivotal, the more responsibility participants assigned to that member. Further, normality considerations played a key part in people's judgments. People assigned more responsibility when a committee member's vote was in disagreement with how most of the other members voted. Participants were also influenced by party affiliation: committee members who voted against what was to be expected given their political affiliation were judged more responsible.

Criticality by itself was not a significant predictor of people's judgments. However, the results suggest that criticality considerations might have still influenced people's judgments. The effects of a voter's party affiliation were strongest in situations in which the policy didn't pass. In these situations, both criticality and normality considerations are aligned. Effects of party affiliation were weak for situations in which the policy was passed. Here, criticality and normality considerations pull in opposite directions.

The current work extends our previous work on responsibility attributions in groups by showing how prior expectations influence people's judgments in a systematic way. While most research that has looked into the influence of normality on people's cause and responsibility judgments has relied on vignette studies (e.g., Knobe & Fraser, 2008), the voting paradigm allows us to probe people's intuitions in a large set of situations that help to tease apart the different factors that influence people's judgments. In future work, we will look into situations in which committee members differ with respect to how much power they have over the outcome.

Acknowledgments We thank Mike Pacer and Max Kleiman-Weiner for many valuable comments. TG and JBT were supported by the Center for Brains, Minds & Machines (CBMM), funded by NSF STC award CCF-1231216 and by an ONR grant N00014-13-1-0333. JYH was supported in part by NSF grants IIS-0911036 and CCF-1214844, AFOSR grant FA9550-08-1-0438, ARO grant W911NF-14-1-0017, and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by AFOSR under grant FA9550-12-1-0040.

References

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, *22*(1), 93–115.

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. (2012). When contributions make a difference: Explaining order effects in responsibility attributions. *Psychonomic Bulletin & Review*, *19*(4), 729–736.

Gerstenberg, T., Ullman, T. D., Kleiman-Weiner, M., Lagnado, D. A., & Tenenbaum, J. B. (2014). Wins above replacement: Responsibility attributions as counterfactual replacements. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2263–2268). Austin, TX: Cognitive Science Society.

Goldman, A. I. (1999). Why citizens should vote: A causal responsibility approach. *Social Philosophy and Policy*, *16*(2), 201–217.

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *British Journal for the Philosophy of Science*.

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*, *56*(4), 843–887.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, *11*, 587–612.

Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The cognitive science of morality: intuition and diversity* (Vol. 2). The MIT Press.

Kominsky, J. F., Phillips, J., Knobe, J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal superseding. *Cognition*, *137*, 196–209.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, *47*, 1036–1073.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, England: Cambridge University Press.

Scanlon, T. M. (2009). *Moral dimensions*. Harvard University Press.

Zultan, R., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: Counterfactuals and causality in group attributions. *Cognition*, *125*(3), 429–440.