Responsibility Attribution for Collective Decision Makers

Raymond Duch
Nuffield College Oxford

Wojtek Przepiorka
University of Oxford, Department of Sociology
Nuffield College
Centre for Experimental Social Sciences Oxford
ETH Zurich, Chair of Sociology

Randolph Stevenson
Rice University

Working Paper Nuffield Centre for Experimental Social Sciences, July, 2012

*November 5, 2012

*This work was partly supported by the Swiss National Science Foundation [grant number 100017_124877]
Abstract

We assess the micro-foundations of rational theories of vote choice in contexts with multi-party governing coalitions. These theories of coalition-directed voting presume voters can assign responsibility to specific parties in the governing coalition. We conduct a laboratory and an internet survey experiment designed to tease out the heuristics subjects use in their responsibility attribution decisions. The lab experiment comprises a group dictator game with weighted voting power of the dictators and a punishment possibility for the recipients. Our results show that recipients punish unfair allocations and mainly target the decision maker with agenda power and with the largest vote share. We find rather weak evidence that decision makers with veto power are targeted or that recipients engage in punishment proportional to weighted voting power. The survey experiment tests whether subjects indeed believe that the decision maker with agenda power has the most influence on the collective decision outcome. The results confirm this conjecture.
1 Coalition Directed Voting

The notion of voters forming performance expectations about, or attributing responsibility to, a single decision maker has profoundly shaped the voting behaviour literature. Downs points out that voters are faced with the need to form expectations about the relative influence of different coalition parties on collective policy choice in order to rationally vote on policy proposals in coalitional systems. He is skeptical, however, about the ability of voters to meet these informational and cognitive requirements. Consequently, Downs (1957, Chapter 3) argues that ‘[e]ventually each voter either abstains, votes after cutting off his deliberation at some unpredictable point, or decides it is easier just to vote for his favourite party.’ Whereas future-oriented policy voting should be important in non-coalitional systems, it will be muted in coalition contexts because voters can not be expected to meet the informational demands of attributing responsibility in these systems. A similar claim is made regarding retrospective policy voting. Voters in systems in which policymaking power is shared will tend to base their votes less on retrospective performance than will voters in systems in which power is more concentrated (Powell and Whitten, 1993; Lewis-Beck, 1998).

Recent work demonstrates that voters in these coalition systems behave in exactly the way Downs says they cannot – that is significant numbers of voters in coalitional contexts engage in “coalition directed voting”, i.e. tactical voting for particular parties in order to try to bring a preferred coalition to power. Kedar (2005) or Bargsted and Kedar (2009), for example, find that voters in contexts with coalition governments engage in compensational voting, i.e. certain voters vote for more extreme parties with the goal of shifting the policy position of governing coalitions closer to their ideal points. Based on data from 86 election surveys conducted in 22 countries, Duch, May and Armstrong (2010) find that in 75% of these surveys more than 50% of voters make coalition-directed calculations.
Hence there is growing evidence that voters know the number of seats won by parties and the cabinet positions they hold. But how do voters map this observed distribution of responsibility into actual administrative responsibility within the cabinet? Our answer is that voters approximate the fully rational calculus of a coalition-directed vote by employing responsibility attribution heuristics for individual parties making up the governing coalition. This essay identifies these responsibility attribution heuristics. By heuristics we mean strategies that ‘guide information search and modify problem representations to facilitate solutions’ (Goldstein and Gigerenzer, 2002). Heuristics are used when information acquisition is costly and decision making is cognitively challenging (Kahneman, 2011; Simon, 1955) – conditions clearly met in complex coalition systems. Models of heuristic-based decision making incorporate two key components: information acquisition and decision making (or the processing of information) (Gigerenzer and Todd, 1999). The task of assigning administrative responsibility requires voters to first acquire information about the composition of a coalition government and second to map observed characteristics of the governing coalition into shares of administrative responsibility.

With respect to the first requirement, Armstrong and Duch (2010) argue that this kind of information is not overly costly for voters to collect. They contend that the historical patterns of cabinet formation are relatively stable and hence voters can cheaply learn which coalitions are likely to form or not. In addition, media coverage gives voters easy access to information about both party sizes and who is in the cabinet. However, these information assumptions are not a focus of this essay. Here we focus on the second question. What heuristics do voters use to map observable characteristics of the governing coalition into beliefs about the unobservable distribution of responsibility?

We use experiments to isolate the specific heuristics for responsibility attribution favoured by voters who only observe collective decisions made in majority rule situations.

1Others have made similar arguments (Barnes, Duch and Stevenson, 2010; Stevenson and Vonnahme, 2011)
Four heuristics are considered in this essay: voters attribute responsibility to decision makers with agenda settings power (i.e., proposer power); they assign responsibility to the largest decision maker; responsibility is attributed proportional to the voting weights of decision makers; and voters attribute responsibility to decision makers with veto power.

The identity of the decision maker with positive agenda power is one of the pieces of information that is frequently communicated to voters. In collective decision making contexts, decision makers with agenda, or proposal, power, irrespective of their voting weight, have been shown to exercise considerable influence on collective decisions (Cox, 2006). With respect to coalition governance there is evidence that Ministers, because they have agenda power related to their portfolios, strongly influence policy outcomes within coalition governments (Laver and Shepsle, 1996).

Secondly, one might expect the largest party in a governing coalition to dominate the public debate concerning a particular coalition policy proposal. And the decision maker with the largest weight frequently shares other characteristics that are likely correlated with influence over collective decisions; such as agenda setting, and veto, power. Hence, a “fast and frugal” heuristic could be the attribution of responsibility to the largest decision maker.

A third heuristic is the voting weights of the individual decision makers – these may be associated with shares of responsibility for collective decisions. There have been many different attempts to develop algorithms that map voting weights into voting power (Banzhaf, 1965; Shapley and Shubik, 1954; Straffin, 1978; Gelman, Katz and Tuerlinckx, 2002; Gelman, Katz and Bafumi, 2004; Heard and Swartz, 1999). The most consistent message one can glean from this large literature is that there is no simple one-to-one mapping between a decision maker’s seat weight and her influence on decisions (i.e., the relationship is not proportional).

Nevertheless, there is evidence that voters associate voting weights of coalition mem-
bers with voting power or influence on coalition decisions (Anderson, 1995; Duch and Stevenson, 2008; Kedar, 2009). This notion that influence over coalition outcomes should reflect the proportion of seats allocated to coalition parties reflects Gamson’s “Law of Proportionality” Gamson (1961, 382) and is widely accepted as the basis for bargaining amongst coalition partners for portfolio allocations (Bueno de Mesquita, 1979; Browne and Frendreis, 1980). Hence, its certainly a plausible heuristics although, again, one for which the empirical evidence is ambiguous.

Finally, a fourth plausible attribution heuristic is the majority, or veto, status of the decision maker in the collective decision entity. If a decision maker has a veto, this heuristic suggests he has significant influence on the outcome. Formal theories of collective choice certainly show that the veto – or the threat of a veto – will affect collective decision making outcomes (Cox, 2006; Cox and McCubbins, 2001; Tsebelis, 2002). And there is some experimental evidence suggesting that veto players can significantly impact the outcomes of collective decisions (Wilson and Herzberg, 1987).

In our experiments we restrict ourselves to exploring how this finite set of heuristics informs responsibility attribution. Subjects could rely primarily on positive agenda power; focus on the decision makers with the largest voting weight; attribute responsibility according to voting weights; or favour a negative agenda power heuristic. The experimental approach we adopt here allows us to disentangle the effect of, for instance, agenda setting power versus veto power in a very controled setting, which is difficult to achieve based on observational data. Only having confirmed the nature of this heuristic reasoning in the lab will we then be prepared to address external validity (Morton and Williams, 2009, 253).
2 Punishment in a Collective Dictator Game

We argue that many collective decisions, including coalition government policy making, resemble other-regarding games such as the dictator game. In a standard dictator game, the decision maker is an individual, but there is a body of experimental research on other-regarding games in which decisions are made by a group of decision makers rather than an individual (Engel, 2010). Our interest here is how the average recipient, in a punishment version of these games, attributes responsibility to the individual decision makers responsible for the collective decision. To our knowledge no experimental evidence has directly addressed this issue.

Our first experiment implements a collective dictator game with the possibility for recipients to punish the decision makers (DMs). Our game has the following characteristics: 1) a decision is taken by a group of five DMs; 2) DMs have varying voting weights and one DM is randomly assigned to have proposal power; 3) the collective decision is taken by weighted majority voting; 4) the decision affects the pay-offs of the recipients; and 5) recipients have the opportunity to punish the DMs after observing the collective allocations. Our primary interest is whether the punishment decisions in our experiment provide insights into the heuristics that individuals employ for attributing responsibility for individuals within a collective decision making entity. We exploit the extensive findings indicating that dictators keep a larger share for themselves and the fact that, given the opportunity, recipients will punish dictators for their unfair allocations (Fehr and Gachter, 2000; Fehr and Fishbacher, 2004).

2.1 Experimental Procedure and Design

Our first experiment was conducted with 71 student and non-student participants at CESS (University of Oxford, Nuffield Centre for Experimental Social Sciences) in March
2011. Each experimental session comprised 20 rounds. In each round, five DMs were chosen at random and they were given as many pounds as there were participants in the experimental session (e.g. £23 if there were 23 participants). The recipients received no endowment. There were four voting weight distribution treatments: (53%, 29%, 10%, 6%, 2%); (48%, 19%, 14%, 11%, 8%); (38%, 21%, 17%, 13%, 11%); (23%, 21%, 20%, 19%, 17%). The treatments were randomised across rounds so that, in each round, one of the four voting weight distributions was randomly chosen and each DM was randomly assigned a voting weight from this distribution. Moreover, one of the DMs was chosen at random to propose an allocation of the money between the DMs and the recipients. In each round, 18 or 20 recipients evaluated each of the five DMs with respect to the collective allocation decision. This gives a total of 5,600 evaluation decisions.

The proposing DM could choose any division of the endowment between the DMs and recipients in units of £1. Next, each of the five DMs cast all their votes for the allocation or all against. Abstention was not allowed. If the allocation received at least 51 votes then it passed. Otherwise the proposing DM had to propose a different allocation, which was also voted on. No other communication was allowed. If three consecutive proposals failed to obtain 51 votes, the group was disbanded and no one was paid anything for that round. In case of a successful proposal, the amounts allocated were split equally among the members of each group. For instance, if the proposing DM decided to allocate £10 to the DMs and £13 to the recipients, then each DM earned £2 and each recipient earned £0.72, if there were 18 recipients.

Recipients were informed of the allocation decisions. In case of a successful proposal this included the amounts allocated to the decision-making group and to the recipients and that this allocation had received a majority of votes. Subjects were always shown the distribution of voting weights over DMs but were never informed of the voting decisions.

2A more detailed description of the experiment is available in the Online Appendix.
of individual DMs. Then, recipients reviewed the allocation and decided whether (and how much) to decrease the income of each DM. In each round, each recipient was given 30 “deduction points” that he or she could keep or assign to individual members of the decision-making group. Recipients could assign all 30 points to one DM, six to each, no deduction points to any DM, or anything in between. Each recipient decided independently about the assignment of deduction points. A DM’s deduction for the round was the average of the deduction points assigned to him or her by all recipients times £0.1. For example, if 10 recipients assign 15 deduction points to a DM, five assign ten points and five zero points, then the assigned deduction was £1 (\(£0.1 \times (10 \times 15 + 5 \times 10 + 5 \times 0) / 20 = £1\)). A DM’s deduction was subtracted from his or her payoffs from the round. However, after each round, the DMs were given no feedback on the amount of their deduction – they only learned their total payoff at the end of the experiment. If the allocation was unsuccessful, recipients were only informed that the allocation failed to receive a majority of votes.

2.2 Results

2.2.1 Punishment Levels

We first examine whether this experimental set up actually produces behaviour that looks like performance voting. Figure 1 presents results for the three sessions in which the proposing DM was known and shows that there is a strong correlation between the amount the DMs collectively chose to keep for themselves and the total level of punishment (i.e., summed across all DMs). The size of the circles is proportional to the number of rounds in which a given amount was kept by the DMs. The lowest observed amount that DMs keep for themselves is £5 which corresponds to the lowest average deduction points which

\[^3\text{We also had an information treatment in which subjects did not know the identity of the proposer. The results reported in the text are only for those treatments in which the proposer is known. The Online Appendix discusses the results of these treatments.}\]
are in the range of 5 points. The maximum amount DMs kept is £25 (nothing for the recipients) which corresponds to the highest average deduction points. This result is consistent with findings from other dictator game experiments with punishment (Bering, 2008; Henrich et al., 2006; Fehr and Fishbacher, 2004) and gives us confidence that our experimental set-up captures the fundamentals of performance voting, thus providing an appropriate vehicle for exploring the way subjects distribute punishments across DMs.

Figure 1: Deduction points assigned by amount kept by DMs

![Graph showing the relationship between amount kept by DMs and average level of punishment.](image)

2.2.2 Heuristics

We begin by identifying the punishment heuristics that subjects appeared to use – there are essentially six:

- “No punishment” – Subjects allocated no deduction points.
- “Punish all DMs equally” – Subjects allocated > 0 deduction points and gave equal shares to all DMs. In the vast majority of cases, they allocated all 30 of their
available points (so six points to each DM).

- “Punish only proposer” – Subjects only allocated deduction points to the proposer. In 85% of such cases they allocated all (30) of their points to the proposer. In 12% of the cases, the proposer was also the DM with the largest voting weight; in 5% of the cases, he also had a majority of the vote weight.

- “Punish proposer and DM with most Weight” – The subject split her deduction between the proposer and the DM with the largest vote weight (which was not the proposer in these cases). No other DM was punished.

- “Punish equally, but more to proposer.” – Subjects punished the proposer more than other DMs but otherwise punished all other DMs equally (and > 0).

- “Punish proportional to vote weight, except for proposer” – Subjects did not fall into any other category and punished DMs in a way that resulted in a positive relationship between vote weight and their punishment, controlling for their allocation to the proposer which was disproportionately large.

Punishment heuristics heavily emphasised the proposer. Subjects paid much less attention to the relative vote weights of the DMs. Indeed, in only 24% of the cases overall did subjects use a punishment heuristic that depended at all on the vote weights of the DMs and in about half of these the only vote weight information that was used was the identity of the largest party.

The experimental results clearly suggest that as allocations become selfish (so performance is judged negatively), the majority of subjects appear to blame the proposer; even when this proposer is not the largest DM. Figure 2 consolidates the heuristics above into those with generous or “fair” allocations (1-9 to the DMs), mixed (10-14 to the DMs) and selfish or “unfair” allocations (15 or more to the DMs). Figure 2a makes it clear that
proposer-directed punishment replaces none-punishment as the preferred heuristic when we move from fair to unfair allocations.

Figure 2b summarises the information a subject employs in each punishment heuristic. Clearly, as allocations become less fair, subjects use more of the available information overall. However, as they become more unfair, they focus less on making distinctions about the sizes of all DMs – only about 20% employ this punishment heuristic in the unfair allocation cases. In most unfair cases (80%), subjects focus exclusively on punishing the proposer without considering the sizes of DMs. Finally, we have cases in which some information about DM size is used (proportional allocations and special punishment of the largest parties). There are no cases in which size information is used and no information about the identity of the proposer is used. In all cases in which size information is used, subjects also accounted for the proposer (or the two were indistinguishable, i.e., when the proposer was also the largest DM).
Figure 2: Punishment Heuristics and Fairness

(a) Heuristics by Fairness

(b) Information Content of Heuristics
2.2.3 Multivariate Model

Figure 2 quite compellingly suggests that responsibility attribution is focused on DMs with proposal power rather than on the voting weights of DMs. Accordingly, we estimate a series of statistical models in which we model how much our subjects punished DMs with different characteristics, while controlling for the amount the DMs as a group kept for themselves. We include a full set of indicator variables for DMs’ voting weights (we impose no limitations on the possible function form between voting weights and punishment) and we include a dummy variable for whether the particular DM was the proposer. Note that the indicator for the largest DM heuristic will be the indicator for the largest voting weight in each distribution (i.e., 23, 38, 48, or 52).

Each row of the raw data records the number of deduction points (out of the 30 possible) that a given subject allocated to a given DM in a given experimental period (where we also record the characteristics of the DM, like voting weight and whether that DM was the proposer in that round). Since there are five DMs in each period, we have five allocations of deduction points from each recipient in each period. In order to identify heuristics about responsibility attribution we focus on the share of the total deduction points that the subject allocates to each DM. Such shares will sum to one for each subject who allocated at least one deduction point to one of the five DMs. Hence, our dependent variable for each subject is a five-part composition that gives the share of the total deduction points allocated to each DM.

We follow Atchisons recommendations (Aitchison, 1986) in estimating this as compositional data. First, we choose a baseline category (i.e., allocations to a specific DM) and then divide each of the other four allocations by this baseline. We then take the log of each ratio. Our dependent variable, in a Seemingly Unrelated Regression (SUR), is the vector of the log of these four ratios. Once we obtain estimates of the parameters of this model, we then get predicted shares of punishment for all five DMs (for a given
subject-period) by transforming back to the original shares via the multivariate logistic transformation (Aitchison, 1986). Finally, we add confidence intervals around our predictions and substantive effects using the simulation methods described in Katz and King (1999). The correlation across DMs for a given recipient is accounted for in the estimation of the SUR (which estimates the correlation across equations). To address the possible correlation across periods for the same subject, we also estimated models with robust standard errors, clustered on subject.4

Since we cannot characterise shares of punishment over DMs when recipients punished no one, in the compositional analysis reported below, we use only those subject-periods in which some allocation of deduction points was made (35% of recipients allocated none of their deduction points while 55% allocated all of them). Thus, the results from these models should be interpreted as conditional on recipients in a period making some allocation. Of course, one can still model the initial decision to punish or not; but, as it turns out, this analysis (included in the Online Appendix) reveals one overwhelmingly significant predictor of non-punishment: the generosity of the collective allocation from the DMs to the subjects.5

We estimate a seemingly unrelated regression that includes four equations (one for each log-ratio with the DM with the largest vote weight, DM 5, as the baseline category). Each equation includes indicators for the vote weight variables associated with each distribution. Also included are indicators of the proposer status of all the DMs. We also include a control in each equation for the total number of deduction points allocated in the period by the subject. This allows for the possibility that subjects change the

4In order to produce these estimates we first replicated the jointly estimated SUR models by estimating each equation separately. In this application, these coefficients are identical, though there are very small changes in the standard errors. Next, we clustered on subject in the separately estimated equations. This produced larger, but still very small differences in standard errors. These results are provided in the Online Appendix.

5Most recipients who punish do not punish all DMs. Consequently, some of the shares in each recipient-periods five-part composition are usually zero. Following Aitchison (1986), we set zero shares (given they are not all zero) to a very small number (.00001) and then rescale to assure the composition still sums to one exactly).
distribution of their punishments as they punish more or less in total. Finally, a variable is included in the model that controls for the allocation decision by the DMs (the split between DMs and recipients). Since interpreting coefficients from compositional models with several sets of indicator variables can be difficult, we focus here on presenting the substantive implications of the model. The full set of estimates on which these substantive effects are based is included in the Online Appendix.

Figure 3 gives a substantive picture of the results for a case in which the DMs’ collective choice was to keep 20 pounds for themselves and a subject’s total punishment was 30 points. The circles with the light grey confidence bands (again, using the simulation methods described in Katz and King (1999)) indicate the estimated share of punishment allocated to proposing DMs with the corresponding voting weight on the x-axis. The circles with the dark confidence bands are estimates of the average punishment allocated to non-proposing DMs, with weights corresponding to the x-axis, (these are averaged over different proposer treatments for this distribution). For example, in the upper-left distribution (2, 6, 10, 29 53), when the DM with a weight of 2 is the proposer, recipients allocate about 20 percent of their deduction points to this DM. This compares to the circle at essentially zero which is the average estimated share of deduction points allocated to non-proposing DMs with a weight of 2. Still focusing on this upper left distribution, note that the estimated deduction points for a DM proposer with vote weight 53 is very similar to that of the average estimated deduction points for non-proposers with this weight – both approach 100%.

There are three results that stand out in Figure 3. First, a proposing DM gets punished

---

Substantive results calculated at other allocations give the same message, though consistent with the descriptive results reviewed in the last section, the impact of the proposer and largest parties relative to more equal allocations declines as DMs are more generous. In the Online Appendix, we provide similar graphs that give the estimates separately for cases in which each other DM was the proposer (e.g., separate estimates for the predicted share of punishment for a DM with vote weight 6 in distribution 1 when each of the other DMs - with weights 2, 10, 29, and 53 respectively - was the proposer). Those graphs also provide distributions of the raw data plotted alongside the relevant estimates and so give a rough indication of how the estimates stem from the uncontrolled data.
more than non-proposing DMs with the same vote weight. Though the simulated 95% confidence bands sometimes overlap, in all cases the point estimates for proposers and non-proposers with the same vote weight are statistically different from one another. Thus, consistent with our analysis of these data in the previous section, subjects use proposal power as an important guide to attributing responsibility to DMs. Secondly, the plurality DM receives more punishment and punishment increases as the plurality DM differs in size from the other DMs. Finally, the results very clearly show that, other than the plurality effect, voting weight is not used by subjects as a general cue for attributing responsibility. There is no clear pattern between voting weight and punishments that applies across the distributions, or even within individual distributions.
Figure 3: The Impact of Voting Weights and Proposal Powers on Punishment

Dark gray bars are estimates for non-proposers and light gray bars are for proposers. The y-axis is the share of a respondent’s total allocated deduction points that was allocated to a DM with the indicated characteristics. The hollow circles are point estimates of predicted punishment shares for a typical respondent (who allocated all 30 of her possible deduction points in a situation in which the DMs kept, a relatively selfish, 20 points of the initial endowment for themselves). These predicted effects are derived from estimated compositional model described in the text (with estimated coefficients given in Table A2 in the Online Appendix). The bars are 95% confidence bands calculated via simulation (Katz and King, 1999). The estimates for non-proposers are averages of the effect for all the occasions when the DM with the indicated vote weight was not a proposer.
Two more subtle results can be discerned in Figure 3. The first is that there appears to be an interaction effect between proposal status and DM size. Specifically, we see that when both the plurality signal and proposal power signal point to the same DM, subjects almost universally punish only this DM (with a predicted share of punishment almost at 100% even in the most egalitarian distribution of vote weights). Finally, our estimates do not suggest that subjects gave any special weight to the fact that in the first distribution the plurality DM could veto any proposal. To see this more clearly, Figure 4 gives the same estimates as in Figure 3 but only for the plurality DM in each distribution. The lower fitted line captures estimated punishment shares that are solely accounted for by the plurality size effect.

Examining the estimates in this way, one can immediately see that our subjects, while they did punish larger DMs more than smaller DMs, did not give an additional increment of punishment (not accounted for by a linear relationship between size of the plurality party and punishment) to DMs having veto power. If this were the case then the estimated punishment for the veto DM (53 voting weight) would have been significantly above the plurality effect fitted line. Had subjects been fully cognisant of the veto DM’s powers, one might have expected them to only punish this DM even when she did not have proposal powers. However, given the point estimate and confidence band (deduction point shares are about 85% when the DM does not have proposal power which is significantly lower than the essentially 100% when the DM has proposal power) this is clearly not the case. Thus, unlike the case of proposal powers, subjects seem less able to recognise the special significance of veto power.

2.2.4 Discussion

Our experiment was designed to assess the importance of the following cues for responsibility attribution: the relative weights of the DMs; the largest DM; the DM with
Figure 4: The Impact of Voting Weights and Proposal Powers on Punishment

Dark gray bars are estimates for non-proposers and light gray bars are for proposers. See note to figure 4 for more details.

veto power over a decision; and the agenda setting DM. Rather surprisingly, with the exception of punishment for the largest DM, this experiment provides little support for the notion that individuals apportion responsibility for collective decisions according to the voting weights of the DMs. The results are also surprising in that the subjects exhibited a clear tendency to hold the agenda setter accountable. Less surprising, but very prevalent, is the tendency for subjects to punish the largest DM. And as the largest DM’s vote gets smaller relative to the other DMs, this punishment declines in a very linear fashion. Indeed, given this linear trend it is not clear in these data that there is a bonus to being above a majority threshold (i.e., a veto player effect) that comes in addition to
the “largest” weight effect. Overall it looks like subjects are looking for clear cues like proposal power and plurality weights. When they find these cues they use them while excluding other information from their responsibility attribution calculations.

One is tempted to conclude from this that subjects in these experiments use these two simple cues because they believe, consistent with academic work on collective decision making, that proposal power and plurality status confer significant influence on collective decisions while vote weight and veto power are less reliable indicators of policy influence. But this is not necessarily the case. An alternative explanation is that individuals seek out a single target for punishment when collective decisions have a negative impact. Punishment may be more of an expression of anger than a rational assessment of responsibility. Emotional psychological models of voting behaviour certainly suggest that the expression of anger in vote choice decisions is associated with an ability to focus blame on a specific target (Ekman, 2003). Perhaps the subjects who are angrier about unfair allocations seek a focus for that anger rather than trying to assess who actually influenced the policy outcome. Our second experiment is designed to explore the extent to which individuals believe agenda setting power actually determines collective decision making outcomes.

3 Internet Survey Experiment

Our second experiment aims to determine whether the proposer-centric punishment identified in the previous experiment results because individuals think that the proposing DM has more influence over the collective decision. The experiment is designed to recover expectations regarding decision making influence. It does not invoke punishment or reward on the part of subjects. This is an Internet experiment based on a sample of 1004
UK respondents that was conducted by the Nuffield CESS July 1-7, 2011.\(^7\)

Respondents were asked to guess the outcome of a collective decision that had been decided by the weighted vote of five DMs prior to the survey. Depending on how closely their guess matched the collective decision, they could earn from 0 to 30 SSI points (more accurate guesses paid more). Each respondent was asked to guess the outcome for three decisions taken under different distributions of voting weights assigned to the five DMs. In the case of each distribution, DM voting weights and DM proposal status were randomly assigned to respondents.

The outcome of the collective decision was described to the respondents with a picture that illustrated the information (these screen shots are included in the Online Appendix). Respondents were told that five DMs had been given a total of £30 and the DMs were asked to decide, based on a majority weighted voting rule, how much should be given to two different charities - an animal shelter and a soup kitchen. The decision making situations differed in three ways. First, in the picture respondents were shown a one-dimensional space that located each DM’s preferred donation amounts for the animal shelter (the balance going to the soup kitchen): £4, £10, £16, £21, £28. Second, the voting weight associated with each of the five DMs was also included in the picture. Three different voting weight distributions were used: [.02, .06, .10, .29, .53]; [.11, .13, .17, .21, .38]; and [.17, .19, .20, .21, .23]. The voting weights in each distribution added to 1 and they determined how much each DM’s vote counted when the donation proposal was voted on. Finally, the picture also indicated which DM was chosen to propose the allocation of the £30.

Respondents provided guesses for outcomes under each of the three distributions. Within a distribution, the particular allocation of weights over the five DMs (i.e., over the five ideal points) that a respondent saw were randomised (there were 120 possible

---

\(^7\)The internet panel sample was provided by Survey Sampling International (SSI) - their panelists are compensated with SSI points which are exchanged for money. A complete description of the survey and sample characteristics is available from the authors.
assignments of weights to positions). In addition, the identity of the proposer that each respondent saw (for each decision) was randomised (so there are 25 possible assignments of a weight and position combination to a proposer in each decision situation). Given the 1,004 respondents, this results in about 40 respondents per proposer/position combination.\(^8\)

### 3.1 Results

The goal of the empirical analysis reported in this section is to explore which characteristics (or combinations of characteristics) of DMs condition respondents beliefs about the outcomes of collective decisions. Building on the results from the lab experiments reported above, we focus on two kinds of characteristics: agenda setting powers – both positive (proposal) and negative (veto) – and the distribution of voting weights (including which DM has the largest weight as well as other more extensive uses of size cues e.g., proportional influence).

To accomplish this goal, we asked each survey respondent to guess the policy choice for three separate collective decisions (corresponding to the three different distributions of voting weights provided above). In addition, the identity of the proposer and the correspondence between the voting weights and policy preferences of the DMs were randomized across respondents. Thus, for each respondent, for each of these questions, we calculated the spatial distance between the respondents guess about the collective decision and the ideal points of each of the five DMs (below, we refer to this variable as distance). This gives us five data points for each respondent for each of the three questions. Taken to-

---

\(^8\)While there are technically 120 \(\times\) 5 cells in our design matrix for each decision (i.e., assignments of weights to positions \(\times\) assignment of proposer) and so a small number of respondents in each cell, this is not consequential for our study since we are only concerned with two aspects of seat weights: (1) whether there is parametric relationship between weights and how close respondents think the policy will be to a given DM, and (2) whether there is a “largest DM” effect. Since we can examine the first of these by fitting parametric models across cells of this design matrix (as we will see in the statistical model given below), the information from a relatively small number of respondents in each cell is easily aggregated to bear on the main questions of interest here.
gether, these data points contain information about the respondent’s beliefs about the relative influence of different DMs (with different seat weights, positions, and agenda powers) had on the collective decision.⁹

Table 1 reports estimates of three separate regressions (one for each question) of our distance variable on measures of the agenda powers, vote weights, and policy preferences of DMs.¹⁰ Specifically, we include an indicator variable for whether the DM was the proposer and expect a negative effect for this variable (e.g., respondents believe proposers are able to move collective policy choices towards their ideal point). We also include indicators for each vote weight in each distribution, leaving out the smallest weight as the baseline category. Thus, we expect (consistent with the results of the lab experiment) that having the largest vote weight will have a significant (negative) impact on distance, but expect no impact (or at least a much smaller and inconsistent impact) for vote weights more generally.

Finally, we include a dummy variable for each DMs policy position (recall that the distribution of policy positions was the same for each question, though vote weights and agenda powers differed across questions and respondents). There are a number of reasons we need to control for policy positions in this analysis. First, we would expect a simple mechanical effect of the centrality of a DMs position on distance. To see why, consider the

⁹We do not assume that a respondent who believes the policy outcome will be near the ideal point of a given DM necessarily thinks that the party is influential. It could be, for example, that the respondent believes the policy will end up being in the middle of the policy space because two relatively influential DMs with opposing and extreme preferences compromise on the policy leaving it close to a more centrally located but non-influential DM. In the empirical analysis we attempt to isolate this sort of incidental influence from beliefs about real influence by focusing on how the addition or removal of DM characteristics (like agenda powers) change respondent beliefs while holding constant the ideological positions of the parties. To do this, we include in the model a flexible set of controls that let us account for all the possible ways in which relative position can impact beliefs (including centrality); thus, our inferences about the impact of other DM characteristics hold for any particular policy preference the DM may hold.

¹⁰Reported standard errors for these regressions are robust estimates clustered on individual respondent. We have also estimated hierarchical linear models that include error components (random effects) at the individual level. Results are quite robust to these alternative estimation strategies, as we would expect since there are not likely to be many unmeasured variables that would cause a systematically larger or smaller distance between a given voters guess and the positions of each of the other parties.
case in which respondents guesses are randomly and uniformly distributed in the policy space. Obviously, there can be no behavioural impact of policy position on distance in this case. There will, however, be a mechanical effect. Specifically, if we calculate the distances between these uniformly distributed guesses and the position of each DM and then take the average of these distances across guesses for any DM, more centrally located DMs will have shorter average distances. In addition to this kind of mechanical effect, it could also be that there is an incidental centrality effect. For example, suppose that respondents use seat weight plurality and proposal power as cues for policy influence and that in a particular treatment, the proposer occupies the left-most policy position, while the plurality DM occupies the right-most position. In this case a respondent might expect a central outcome, not because the non-proposing, smaller, but centrally located DMs are thought to have influence, but because the DMs that do have influence happen to be located at opposing poles. Nevertheless, the distances will be shorter for such non-influential central DMs than for less central plurality and proposing DMs. Thus, both the mechanical effect and this incidental centrality effect suggest that it is crucial for us to control for the spatial location of the DMs in our analysis and leads us to expect that there will be a strong empirical impact of these positions (with distances shorter for central DMs regardless of their seat weights or proposal powers). Finally, it is of course possible that respondents actually use centrality as a cue (like proposal power or seat weights) to assign influence over decisions. Our design, however, is not optimal for isolating this effect from the mechanical and incidental effects described above. Indeed, a survey experiment designed to do so would face significant design challenges since in any unidimensional policy space such effects are impossible to avoid. For our purposes, this limitation is not consequential since the focus of this experiment is to explore the

---

11 A simple example makes this plain: Suppose five DMs (called A, B, C, D, and E) occupy positions 0, 1, 2, 3, and 4, respectively. Now, suppose there are 4 guesses distributed uniformly over the same space (say at positions .5, 1.5, 2.5, and 3.5). Summing up the distances between each guess and each DMs position and taking the average over guesses gives an average distance for each DM as 2, 1.25, 1, 1.25, and 2, respectively.
insights about vote weight and agenda setting cues that came out of the lab experiments (where spatial issues were absent) rather than to induce all the possible cues respondents may be using. Controlling for such influences, however, is quite important and so we do so in our empirical models.

Looking at the coefficients in Table 1, we see some initial evidence that respondents in this experiment are acting much like the subjects in the lab experiment, giving us some confidence that the attributional cues identified in that experiment are really about identifying decision-making influence. Specifically, we see that the coefficient on the proposal dummy is strongly significant in each equation. Further, in the first and second equations, the dummy variable indicating the DM with the plurality seat weight (.53, .38, and .23 in the respective equations) is significantly different from both zero and the coefficients on the dummy variables for the sizes of other (smaller) parties. In the last equation, no such effect is apparent. But given the very egalitarian distribution of vote weights used in this question (i.e., 17, .19, .20, .21, .23) this is really the exception that proves the rule: when size of the “largest” DMs is qualitatively not really different from the other DMs, then being the largest does not provide a very good cue for attributing influence.

Looking at the impact of vote weight more generally, the results are again largely consistent with our lab experiment. Indeed, the coefficients reveal that the impact of seat weight on distance (for DMs who are not the plurality DM) is often inconsistent (i.e., coefficients do not get monotonically larger more negative - for larger seat weights) and, with the partial exception of the first equation, insignificant (i.e., they are not all different from zero or from each other). Finally, the impact of policy position shows exactly the relationship we expected with distance being shortest for central DMs and increasing (symmetrically) as one moves to more extreme positions.
Table 1: OLS Regression of Individual Guesses and DM Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Distribution 1</th>
<th>Distribution 2</th>
<th>Distribution 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dummy for policy position 10</td>
<td>-5.43</td>
<td>-5.40</td>
<td>-5.17</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Dummy for policy position 16</td>
<td>-7.72</td>
<td>-7.95</td>
<td>-7.49</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Dummy for policy position 21</td>
<td>-5.32</td>
<td>-5.45</td>
<td>-5.17</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.26)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>Dummy for policy position 28</td>
<td>0.80</td>
<td>0.79</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.30)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>Dummy for vote weight .06, .13, .19</td>
<td>-0.43/1/</td>
<td>0.02/2/</td>
<td>0.17/3/</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Dummy for vote weight .10, .17, .2</td>
<td>-0.27</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.20)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Dummy for vote weight .29, .21, .21</td>
<td>-0.44</td>
<td>-0.33</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.20)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Dummy for vote weight .53, .38, .23</td>
<td>-1.65</td>
<td>-1.02</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.21)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Dummy for Proposer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.17)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Constant</td>
<td>12.44</td>
<td>12.09</td>
<td>11.82</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.19)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Rows of Data</td>
<td>5020</td>
<td>5020</td>
<td>5020</td>
</tr>
<tr>
<td>Respondents</td>
<td>1004</td>
<td>1004</td>
<td>1004</td>
</tr>
<tr>
<td>DMs per respondent</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Note: [1] Test to reject -.43=-.27=-.44=0 Pvalue=0.06;
[2] Test to reject .02=.07=-.33=0 Pvalue=.17;
[3] Test to reject .17=.10=.25 Pvalue=.2.
Standard Errors Clustered on Respondent
Figure 5 summarises the multivariate results for the survey experiment. The results from Table 1 indicate that respondents expected proposer status to influence considerably the policy outcome. To illustrate this the graph presents the distance between respondents’ predicted policy outcome and the policy position of the DM with the indicated characteristics. There are three graphs corresponding to each of the three distribution treatments. Each graph presents the distances between the predicted policy outcomes and the DM policy positions for DMs that were proposers (with darker confidence bars) and for those who were not proposers (with lighter confidence bars). A line has been fitted between these two sets of distances excluding the DM with the most weight.

First note that none of the slopes of the lines graphed are statistically significant from zero. They are not even close. So working from the top line in the first graph (proposers for Distribution 1) to the bottom line in the third graph (non-proposers Distribution 3) the p-values for each line are: .37, .29, .30, .38, .56, and .87. This is also born out by the the variety of tests of the equivalence of the various dummy variables noted earlier. We have no evidence that the distance between the expected policy and the DM’s position gets smaller as voting weight gets bigger.

Then note that the distance between the respondent’s predicted policy outcome and the policy position of the proposing DM is always smaller than it is for the non-proposing DMs. And in most cases the confidence intervals on the predictions do not overlap. Hence it is not surprising that in the three graphs in Figure 5, the fitted line for the proposer is consistently lower than it is for the non-proposer. Our principal finding from the lab experiment is confirmed: responsibility attribution is strongly associated with agenda power.

Being the largest party also sometimes matters. In all three graphs, being the largest appears to matter more when you are also the proposer. This was not obvious in Table 1. But it is clear in Figure 5 which shows the interactions that result from the non-linear
model. The fitted-line indicates how much closer the expected policy is to the largest DM than it would be if one projected only a size effect consistent with some kind of proportional influence. In all cases (except for one proposing in the very equal Distribution 3) the point estimate for the largest DM is below the line. Hence the largest party heuristic, for which we do have evidence, is quite distinct from a voting weight heuristic, for which there is limited evidence.

But being the veto player does not appear to matter. The point estimates for the majority DM (for both proposing and non-proposing) is not really different from those for the large but not majority DM (who has .38 vote weight). So there is no real majority or negative veto effect apparent here when one accounts for being the largest.

And, finally, the largest party effect diminishes quite a lot when the “largest” is only marginally larger than the others. In the most equal case (Distribution 3), the dummy for the largest seat weight is not different from zero; and is not different than the other dummy variable coefficients (none of which are different from zero). So in this equal distribution case there is no largest DM effect.
Figure 5: Predicted Distances between Respondent’s Guess and DM Position

- Distribution = {.02 .06 .10 .29 .53}
- Distribution = {.11 .13 .17 .21 .38}
- Distribution = {.17 .19 .20 .21 .23}

Lighter confidence bars are for non-proposers; Darker bars are for proposers.

Fitted lines are calculated using only cases other than the DM with the most voting weight. None of the slopes are statistically significant.
3.2 Discussion

The lab experiment was designed to recover responsibility attribution heuristics subjects employ to punish individual DMs who take a collective decision that affects them. The online experiment, with a representative sample of the UK population, was designed to test whether the punishment heuristics recovered in the lab experiment result from an assessment of the relative importance of individual DMs for the collective decision. Accordingly, in this online experiment, subjects were asked to anticipate collective decision outcomes under a number of different treatments that varied the characteristics of the DMs. The results confirm the two negative results of the lab experiment that voting weights and negative agenda power are of limited importance for responsibility attribution. And the online results confirm the two positive lab experiment findings: respondents are inclined to attribute responsibility for collective decisions to the DM with agenda power and in some cases to the largest party.

As we point out in this essay, there are a variety of heuristics that might shape responsibility attribution but we have little insight into which of these are most likely to be adopted by the average person in the population. The experimental evidence presented here provides a rather stark and quite definitive answer to this heuristic question and it challenges conventional wisdom. Many works in the coalition government literature assume that voters attribute responsibility according to the parties’ relative voting weights in the governing coalition (Kedar, 2009; Duch and Stevenson, 2008). Our results suggest this heuristic is employed with limited frequency. Moreover, individuals do not seem to favour the heuristic that assigns responsibility to the decision maker with negative agenda power. On balance, these heuristics play a minor role in the attribution of individual responsibility for collective decision making.

On the other hand, individuals clearly favour agenda power as a heuristic for attributing responsibility for members of a collective decision making body. Two quite different
responsible attribution experiments produce this same result. The Nuffield CESS lab experiment demonstrates that if individuals have the opportunity to hold individual decision makers responsible for a group decision, they primarily attribute responsibility to the proposer and the party with the largest weight. The second CESS online experiment was designed to determine whether individuals hold proposers responsible because they believe they actually have a disproportionate impact on the outcome (rather than simply constituting a focal point for an individual’s anger or enthusiasm). Results from this second experiment confirm the two lab findings: that individuals believe that the proposer and the largest party in fact have a disproportionate impact on the outcome of collective decision making.

Does this positive agenda power result, identified in the controlled experimental setting, help us explain responsibility attribution in the general voting population? Two empirical issues are of particular interest: Are voters who recognise the importance of proposal power more likely to hold policy makers accountable? And, secondly, are those who recognise the importance of proposal power more likely to attribute responsibility to the individual party in the coalition with proposal power? In fact, in a follow-up study based on the online survey experiment data, the authors (Duch and Stevenson, 2012) model the economic vote and establish that it exhibits both of these characteristics.\textsuperscript{12} Only those individuals who understood the value of proposal power in the experiment exercised an economic vote (i.e., rewarded or punished the parties based on their evaluation of the economy). And second, those who recognised the importance of proposal power focused all of their economic vote on the Conservative Party; the member of the current Conservative-Lib-Dem governing coalition with clear proposal power for economic policy. This finding suggests that the attribution behaviour we identify in the lab maps very nicely onto actual voting behaviour in the general population and hence strengthens

\textsuperscript{12}That paper uses the data from the online experiment described above. The authors use a random coefficient model to estimate an individual-specific effect of proposal power on responsibility attributions (which was possible because each individual made three attributions of policy responsibility).
the external validity of our results.

Our results suggest that future research should design studies of vote choice that explicitly consider how the media frames the agenda power of different parties within the governing coalition; how voters perceive agenda power within the governing coalition; and how these factors in turn condition vote preferences for each governing party. Hence, an important challenge for students of comparative electoral behaviour is to better understand the voter’s beliefs about the roles of parties in the coalition policy making process – in particular who is perceived to exercise positive agenda power – and how this in turn conditions their vote choice.
References


Barnes, Tiffany, Raymond Duch and Randy Stevenson. 2010. Clarity about Clarity: Sorting through Competing Explanations for the Clarity of Responsibility Hypothesis. Paper presented at the 2010 Annual Meeting of the Midwest Political Science Association.


