A Few Bad Apples

A model for moral bankruptcy in representative bodies

Eszter Kiss*

December 1, 2025

Latest version available here[1]

Abstract

This paper explores the dynamics of allocation choices within a team dictator game to model the decision making process within representative bodies. It seeks to answer the question of how elected bodies can become morally bankrupt over time. The models incorporate probabilistic dictator choices to account for negotiation process outcomes and individual susceptibility to surrounding opinions that a deterministic utility maximizing choice does not necessarily account for. Using distributional assumptions on choice probabilities, this paper links convergence of morals within a group to individual morality and susceptibility to group members' opinions.

Keywords: Team dictator games, group morality, random utility

JEL Codes: C73, D70

*University of Chicago

[1]https://github.com/eszter-kiss/Works/blob/main/pdfs/Writing_Sample_Version_Latest_2025.pdf

1 Introduction

This paper seeks to explain how it often seems to be the case that the politicians we elect do not represent our interests, make choices that the collective does not want, and in general, seem to become morally corrupt with time.

Individuals in society have different moral codes and views on ethics and therefore behave heterogeneously during social decision making. Some seem to be less selfish or altruistic than others. Yet this alone does not explain why it seems to be the rule not the exception that individuals become morally bankrupt in representative bodies - unless we believe that voters are consistently fooled by candidates during elections about their true character. It could instead be true that most candidates are not fooling voters regarding their social preferences. Rather, it is the case that people's morality consists of two components - their social norm function (or equivalently, the weight they place on it when maximizing their utility), and a susceptibility to other individuals' moral views. And while we might be able to observe the type of morals an individual holds, we cannot observe the strength of adherence to it.

This paper considers the choices of individuals to be random rather than deterministic. Assuming non-degenerate choice probabilities over the set of possible allocations, the subsequent models introduce a team dictator game within a representative body where final group allocation choices are affected by not only the mean value of individual choice probability distributions but their variance as well. In a multi-period setup, this paper presents a way in which individual moral tendencies converge within a group.

Section 2 surveys the literature on other-regarding preferences and empirical research on team dictator games to support the theoretical results of the model. Section 3 first introduces the deterministic foundation of the main model, then summarizes the literature on random utility theory. Finally, it contains the main models and conclusions of the paper. Section 4 expands upon future areas of research and Section 5 concludes.

2 Models of Altruism

G. S. Becker (1974) coined the term *social income* to describe the sum total of "a person's income and the monetary value of the relevant characteristics of others". His paper discussed a production function where an individual's output of commodities depends on other people (contained within the group of environmental variables). An example of such variables was the opinions of others about the individual. Becker's stated innovation was assuming that the individual can change the values of these variables as opposed to taking them as given and choosing actions accordingly.

Later papers modeling other-regarding preferences instead were inspired by the body of empirical research on behavioral games, including laboratory experiments on the ultimatum and dictator games that started in the early 1980s. The dictator game was designed to assess the level of altruism that individuals exhibit. In the game there are two individuals, the dictator and the receiver. The dictator is given a fixed amount of money to distribute freely among himself and the receiver. There are no game theoretic considerations for the dictator to make and so classical microeconomic theory tells us that any rational, profit-maximizing decision maker will decide to keep all of the money and allocate nothing to the receiver. Consistent results contradictory to this assumed behavior started the rich theoretical literature on other-regarding preferences, altruism, inequity-aversion, reciprocity and related concepts.

Rabin (1993) modeled the "battle of the sexes" game by incorporating a *kindness function* into players' utilities. Using beliefs about the other player's moves, this function captures how kind one player thinks the other is planning to be towards her and accordingly chooses her action so as to treat a kind player kindly and a bad player badly, potentially at the cost of trading off her own material returns for perceived "fairness". Kindness in the model was defined in terms of the size and direction of the deviation from an equitable outcome caused by the planned action. Fehr and Schmidt (1999) avoided the inclusion of beliefs about others' choices and set a model in which people are inequity averse: they suffer negative utility when being allocated more or less than what is equitable. (However, they suffer less when this inequity monetarily benefits them as opposed to the other parties.) When simplified to two-player settings, the proposed utility function was presented as

$$U_i(x) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}, \qquad \alpha_i \geq \beta_i \geq 0.$$

Bolton and Ockenfels (2000) set up a model for equity, reciprocity, and competition. Their approach was similar to Fehr and Schmidt (1999) in that a person's utility is only dependent on her own monetary payoff and own relative payoff (i.e., the amount of inequity in the distribution of payoffs) and different in that how she trades off her own monetary and relative payoff is private information. Their model as it relates to dictator games in general form is:

$$U_i \equiv U_i(y_i, \sigma_i) = U_i(c\sigma_i, \sigma_i)$$

where the monetary payoff y_i of person i is equal to i's relative payoff σ_i times the total amount of distributable money c. A concrete functional form of this utility function that they provided as an example is the inspiration for this paper's model and is used in a modified form in equation (1).

$$r_i(c) = \arg\max_{\sigma_i} U_i(c\sigma_i, \sigma_i), \qquad c > 0$$

is the allocation choice of a player in the dictator game and the paper's assumption of a heterogeneous population is presented as the fact that all possible values of r have a non-zero density: $f^r(r|c) > 0, \quad r \in \left[\frac{1}{n}, 1\right]$.

As Falk et al. (2008) demonstrated through laboratory experiments, it is not only equitable allocations that matter to participants, but also fairness intentions. Models that incorporate both the kindness function approach of Rabin (1993) and the inequity aversion approach of Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) are expected to fit the data best.

Indeed, people who want to behave selfishly in dictator games but feel social pressure against doing so may strongly prefer scenarios where their choice set is limited to only unfair actions, precisely because they know that provable selfish intentions matter more to others than forced decisions. Dana et al. (2006) showed through experimentation that a significant number of dictators were willing to sacrifice one dollar to avoid playing the dictator game and be allowed to exit secretly with the remaining nine dollars. This is formalized in Dillenberger and Sadowski (2012) through menu-dependent preferences that exhibit susceptibility to shame. In their paper, the agent first secretly chooses a menu of possible allocations between themselves and a receiver, and in the second stage publicly chooses a specific allocation from said menu. The structure of their maximand,

$$U(A) \coloneqq \max_{\mathbf{a}\in A}[u(\mathbf{a}) - g(\varphi(\mathbf{a}), \max_{\mathbf{b}\in A}\varphi(\mathbf{b}))]$$

contains additively a private utility $u(\cdot)$ and a cost $(g(\cdot))$ for deviating from the maximizing choice of some social norm function $\varphi(\cdot)$. This is also the utility of a given menu A that the dictator can choose in secret before making the public choice allocation a from the menu.

The theoretical contributions cited above were preceded by a large volume of empirical results in behavioral economics regarding ultimatum games and dictator games. Laboratory experiments demonstrated that the assumptions of a purely self-interested player who allocates as little as allowed do not fit the data. Models using these games therefore have to take into account the most well-supported behavioral patterns. A comprehensive meta-study was produced by Engel (2011) on dictator games, with evidence pointing towards 1) identified dictators being more generous than anonymous ones, 2) groups being less generous than individuals, and 3) the presence of vast heterogeneity among humans in terms of generosity and selfishness. Further relevant results were produced by Leider et al. (2009) and Goeree et al. (2010) who showed that social networks and the social distance between dictators and receivers significantly affects the generosity of the dictator.

Group morality has also been a question of experimental research. In an n-person prisoners dilemma experiment conducted by Kerr et al. (2009), members of small groups were individually assigned money to distribute between themselves and the group. Contributions to the group doubled the pooled amount and then were redistributed evenly. Full cooperation led to maximum welfare, but the individual dominant strategy was to keep one's endowment while others contribute. Participants were exposed to other groups with "bad apples"; people who allocated nothing to their own group. Even one "bad apple" in another group significantly induced more selfish decisions in subsequent groups. Cason and Mui (1997) conducted an experiment where participants initially played an individual dictator game, distributing y dollars, and were then placed in teams of two to allocate $2y$ dollars among themselves and two receivers. Results indicated that when participants with significantly differing individual choices formed teams as dictators, other-regarding preferences dominated the allocation. Luhan et al. (2009) replicated this study with dictator groups of three and found the opposite effect. Teams were more selfish during allocation than the average of the individuals comprising the team. After the team dictator game, participants played the individual dictator game again. They were significantly more selfish than their initial play if their team's outcome was more selfish than they initially were. Furthermore, if they were more selfish than their team, their degree of selfishness did not significantly change in the third step of the experiment. In other words, there was a consistent downward convergence to selfishness among participants, and the most selfish players remained more or less the same. Franzen and Pointner (2014) also replicated the 1997 study (with a group size of three) and found no fundamental change in the degree of selfishness of dictators when placed in a group. Final group allocations were driven by the level of individual selfishness that participants initially exhibited. Panchanathan et al. (2013) used an n-player dictator game to evaluate the bystander effect as a function of the size of the group. Group members individually made allocation choices. As the dictator group size grows, the welfare of the receiver can also grow. The bystander effect in this context means that the average allocation decreases with group size. The results showed that the bystander effect materialized. Ito et al. (2016) conducted repeated team dictator games (with dictator groups of three) to assess how the degree of selfishness changes as the game progresses in teams compared to individuals. There was a significant decrease in the offered amount to receivers as the rounds progressed. The researchers contended that this was due to the contagion of selfishness. As less self-interested participants observed that the majority opinion was more selfish than they were, these participants also changed their behavior to be more selfish.

Perhaps as a consequence of less consistent results in team dictator games compared to individual dictator games, significantly fewer theoretical models were constructed to explain the experimental findings than in the case of the original dictator game.

Research on team dictator games has shown more varied outcomes than individual dictator games. The surveyed papers at times observed increase, decrease and no significant change both in individual and group levels of selfishness. Some form of convergence to the behavior of more selfish individuals in a group has been more frequently observed than the contrary. Additionally, it has been noted, even beyond dictator games, that "bad apples" negatively impact the overall morality of individuals in a group. Iterations during the game reveal a tendency to converge to more selfish behaviors both at an individual level and within a group. Moreover, moral decisions that are costly to the individual tend to trigger a bystander or free-rider effect. At the same time, a satisfactory model about group morality should be able to account for all of the perceived changes listed above. All these observations are taken into account and aimed to be incorporated into the model proposed in this paper.

3 Model

3.1 The deterministic setup

There are N individuals denoted $i \in \{1, ..., N\}$. We shall call this group Society. Out of these individuals, $k \ll N$ are elected into a representative body that we shall call Congress. The $N-k$ individuals not in elected roles are the Electorate.

Using the terminology of behavioral economics, Congress is in a dictator position within a multi-player dictator game: it gets to distribute 1 unit between Congress and the Electorate (individuals within each subgroup get an equal proportion of what was allocated):

$$\theta_C + \theta_E = 1$$

where θ_C is the portion allocated to Congress and θ_E is the portion allocated to the Electorate.

A socially fair allocation would be such that all individuals in society receive the same share: $\theta_C = \frac{k}{N}$ is allocated to "Congress" and $\theta_E = \frac{N-k}{N}$ is allocated to the Electorate in the **point of fairness**.

Every individual in Congress faces an internal conflict between their self-interest and their sense of fairness: a fully self-interested individual would want to allocate all of the unit to Congress, while a fully altruistic individual would want to settle on the point of fairness. This problem (modified from the above-mentioned example of Bolton and Ockenfels, 2000) and its corresponding solution [2] can be written as follows:

$$\max_{\theta_C} U_i(\theta_C) = \max_{\theta_C} \left[a_i \frac{\theta_C}{k} - \frac{b_i}{2k} \left(\theta_C - \frac{k}{N} \right)^2 \right]$$

$$\theta_C^* = \min\{1, \quad \frac{a_i}{b_i} + \frac{k}{N}\} \qquad \text{if} \quad a_i \geq 0, b_i > 0$$

$$\theta_C^* = 1 \qquad \text{if} \quad a_i > 0, b_i = 0$$

[2] A simple derivation can be found in Appendix A.

where the first solution represents an individual that is to some extent torn between their self-interest and pro-social behavior (provided $\frac{a}{b}$ is not too large) , and the second solution represents the perfectly self-interested individual. The $\frac{a}{b}$ ratio can be interpreted as the measure of selfishness of the individual since this is the size of perturbation from the point of fairness.

Since utility functions are of ordinal interpretation, $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ describe the same preference ordering even when $a_1 \neq a_2, b_1 \neq b_2$ (one parametrization is a strictly increasing transform of the other). This means that for (partial) preference identification we can reparametrize the problem to a single $\lambda_i \in [0, 1]$:

$$\max_{\theta_C} U_i(\theta_C) = \max_{\theta_C} \left[\lambda_i \frac{\theta_C}{k} - \frac{(1-\lambda_i)}{2k} \left(\theta_C - \frac{k}{N} \right)^2 \right] \tag{1}$$

$$\theta_C^* = \min\{1, \quad \frac{\lambda_i}{(1-\lambda_i)} + \frac{k}{N}\} \qquad \text{if} \quad 1 > \lambda_i > 0 \tag{2}$$

$$\theta_C^* = 1 \qquad \text{if} \quad \lambda_i = 1. \tag{3}$$

which gives us the same set of solutions as any value $\frac{a}{b}$ when a, b are non-negative because $\frac{\lambda_i}{(1-\lambda_i)}$ on $[0, 1)$ is strictly increasing and its range is the non-negative reals. Therefore, it perfectly captures the full possible array of selfishness within a population from completely altruistic ($\lambda = 0$) to completely selfish ($\lambda = 1$).

The above shows the solution to an individual's allocation preference in Congress, but how a representative body itself makes decisions is yet to be discussed. Two key elements have to be taken into account: 1) the preference aggregation of individual members of Congress and 2) the process of negotiations between members of Congress whereby individuals potentially move away from their ideal preference θ_C^*. This model takes a probabilistic approach that will integrate both of these concerns.

3.2 The probabilistic approach to allocation preferences

If an individual were to play the dictator game multiple times over the course of different circumstances, it could reasonably be assumed that he would not make the same allocation choices each time, and such has been confirmed through experiments (Agranov and Ortoleva, 2017). Rather than viewing an individual's (member of Congress) decision as the outcome of a deterministic maximization problem like (1), this paper considers it as a random process where the choice of the agent is random (either because we - or in this context the Electorate - cannot observe every input of his true utility or because he draws his utility randomly from a set before each choice). Here, an individual politician's utility maximizing allocation has a probability distribution over possible allocations, drawn from in each game iteration.

This paper contends that allocation preferences of individuals in a group comprise two components: a deterministic preference function (encompassing self-payoff and fairness weights) and random susceptibility to others' preferences. Hence, an individual's probability distribution over all possible allocation choices both presents their preference (mean, median or mode) and adherence strength to own preference (standard deviation or similar spread metric).

This approach captures how voters might favor a politician whose mean aligns less with their beliefs but exhibits strong adherence over a politician with greater decision variance. In this context, two distributions may have means at different distances from the point of fairness, but if the one with a mean further away has a tighter distribution than the other, the probability of a draw from the former being some fixed, large distance away from the point of fairness may still be smaller than in the case of the latter.

Additionally, this model provides a convenient way of characterizing collective decision-making. Congress's overall preference could be the mean of the individual choices, or a draw from a mixture distribution, with each member's probability distribution as a weighted component.

3.2.1 Stochastic preferences

The theory of stochastic preferences emerged from empirical observations indicating that individuals make different choices when faced with the same decision multiple times. Three primary modeling directions were developed: random utility, bounded rationality, and deliberate randomization (Agranov and Ortoleva, 2017). This paper focuses on stochastic choices within the framework of random utility.

A random utility model is a stochastic model such that from any offered set M (where $M \subset S$ and S is the finite set of all possible choices $1, ..., n$) the individual has a strict preference ordering over all elements of M through a vector of random utilities (G. M. Becker et al., 1963). For n number of choices there exists a $(U_1, ..., U_n)$ random vector such that for all $M \subset S$ $x(M) = Pr[\text{agent chooses } x \in M] = Pr[U_x \geq U_y \quad \forall y \in M]$. When applying this to S, $x(S)$ provides a **choice probability distribution** over all possible outcomes.

The source of randomness in the utilities U_x can be modeled in multiple ways. In G. M. Becker et al. (1963) the randomness of the utilities stems from the fact that each choice $x \in S$ has an array of attributes $\mathcal{A}_x$ of which the agent only considers a random subset $A_x \subset \mathcal{A}_x$ with probability $Pr[A_x]$ when calculating the utilities $U_x(A_x)$. Gul and Pesendorfer (2006) formulate the random utility function as a probability measure μ over a set of possible utility functions $\mathcal{U}$ with deterministic utility elements $u \in \mathcal{U} : S \to \mathbb{R}$. The choice probability $x(\cdot)$ maximizes the random utility function μ if for all $M \subset S$ subsets $x(M) = Pr[u \in \mathcal{U} : \max_{y \in M} u = x]$ where the probability is taken with respect to the μ probability measure. In the **additive random utility** model (Malmberg and Hössjer, 2018)

$$U_i = f(X_i) + \epsilon_i, \quad i = 1, ..., n$$

where U_i is the random utility value of choice i, X_i is a (possibly random) vector of characteristics for choice i, and ϵ_i are independent, identically distributed random variables. When X_i are non-random, the utility function consists of a deterministic component $f(x_i)$ and a random component ϵ_i. The theoretical justification for this is as follows: the random component stems from informational asymmetry where the modeler does not know the agent's exact utility function and the unknown components manifest in an additive random component. In our setting of representative bodies, this amounts to the voters being only partially aware of the factors that an elected official bases decisions on. If $\mathbb{E}[\epsilon_i] = 0$, then this model can be applied to what has been discussed above: the voter knows a candidate's morals (aware of $f(x)$ and the solution to the maximization problem of the deterministic component or the expected utility, $\arg\max f(x)$) but not the strength of adherence to said morals when making decisions in a representative body.

When the set of choices S is finite, research has been produced about setting up axioms that observed choice probabilities have to adhere to in order to be rationalizable by some random vector of utilities $(U_1, ..., U_n)$ consistent with a maximization decision (Falmagne, 1978, D. L. McFadden, 2005, Barberá and Pattanaik, 1986). For $i \in \{1, ..., n\}$ finite choices,

$$V_n = \max_{i \in \{1,...,n\}} U_i$$

is itself a random variable. When ϵ_i are independently identically $EVT1(0,1)$ distributed we get the multinomial logit model (D. McFadden et al., 1974). We are not looking for the distribution of V_n itself (as the actual value of utilities is not of interest due to its purely ordinal interpretation). Rather, we are interested in the probability distribution of

$$I_n = \arg\max_{i \in \{1,...,n\}} U_i.$$

The multinomial logit has an analytical expression for the choice probabilities. But when the choice set is infinite, the random maximization problem becomes theoretically challenging. Research has focused on extending finite cases to infinite cases by taking the number of choices n to infinity and looking at limiting distributions (Malmberg and Hössjer, 2018). In short, the difficulty in the infinite case stems from the fact that the aim is to derive choice probabilities that are consistent with some random utility maximization process, but the assumptions that need to be made on the distribution of ϵ_i are very restricting in regards to what kind of final choice probability distributions we can get (Malmberg, 2013). If the tail distributions are too thin (Gaussian), the choice probabilities will collapse to the choice that maximizes the deterministic portion $f(\cdot)$ of the random utility function. Research in this area is ongoing. Malmberg and Hössjer (2018) set up a model for living location preferences in a disc within $\mathbb{R}^2$ that result in choice probability over commuting distances from work. With an exponential distributional assumption on the stochastic component of the additive random utility, they derive a choice probability distribution where the relative likelihood (maximum of the density function) is largest for the distance that maximizes the deterministic utility of the agents. Even in the standard logit model where $P[\text{choice} = x_i] = \frac{e^{f(x_i)}}{\sum_{j=1}^{n} e^{f(x_j)}}$, the probability is maximized for the choice with the highest deterministic utility component.

Therefore, we can gather the following from existing research: a coherent choice probability distribution on a continuum of choices should be non-degenerate, and its density function should attain its maximum at the choice that maximizes the deterministic utility component. Even though as of now it is unclear whether there is a concrete family of distributions that when assumed of choice probabilities are compatible with utility-maximizing agents, and it is also unclear what family of distributions ϵ_i have to follow for choice probabilities to be non-degenerate and of a tractable analytical form, we proceed with assuming a functional form for choice probabilities and argue that all calculations we make should produce similar results for other distributions with similar attributes.

3.3 The stochastic setup

To reiterate, there are N individuals denoted $i \in \{1, ..., N\}$ Society, with $k \ll N$ elected into Congress, and the $N - k$ individuals not in elected roles are the Electorate. Congress gets to distribute 1 unit between Congress and the Electorate ($\theta_C + \theta_E = 1$). The **point of fairness** is the most socially fair allocation where everyone receives the same share: $\theta_C = \frac{k}{N}$,

$\theta_E = \frac{N-k}{N}$. The full utility function of any candidate for Congress is unknown (random):

$$U_i(\theta_C) = \lambda_i \frac{\theta_C}{k} - \frac{(1-\lambda_i)}{2k}\left(\theta_C - \frac{k}{N}\right)^2 + \epsilon_i(\theta_C) \tag{4}$$

so the Electorate only knows the solution to the maximization of the deterministic part:

$$\max_{\theta_C} u_i(\theta_C) = \max_{\theta_C} \lambda_i \frac{\theta_C}{k} - \frac{(1-\lambda_i)}{2k}\left(\theta_C - \frac{k}{N}\right)^2 \tag{5}$$

$$\theta_C = \begin{cases} \frac{\lambda_i}{(1-\lambda_i)} + \frac{k}{N} & \text{if} \quad 1 - \frac{1}{2-\frac{k}{N}} > \lambda_i > 0^3 \\ 1 & \text{if} \quad \lambda_i \geq 1 - \frac{1}{2-\frac{k}{N}} \end{cases} \tag{6}$$

Due to the randomness or informational asymmetry that the Electorate faces, the actual solution to the maximization problem of a candidate is a random variable X_i called the **allocation choice** of individual i, with range $[0,1]$, and with a probability distribution that we shall call the **choice probability distribution** of the candidate.

Assumption 1 (Non-degenerate choice probabilities)**.** *The shocks $\epsilon_i(\theta_C)$ are such that the choice probability distributions are non-degenerate.*

Assumption 1 is just a restatement of what was discussed in the random utility literature. The sufficient set of requirements on the shocks that guarantee the above assumption to our knowledge have not been laid out in full, but are not integral to this paper with the assumptions below.

Assumption 2 (Choice probability distributions)**.** *X_i are independent across individuals, absolutely continuous random variables with range $[0,1]$, and a unique mode at $\theta_C \in [0,1]$ that maximizes the expected utilities of the individuals.*

Absolute continuity serves as a convenience assumption, and combined with the boundedness of the support (range) of X_i, guarantees the existence and finiteness of the mean and variance for X_i ($\int_0^1 x f_i(x)dx \leq 1$ and similarly for variance). The assumption of a unique mode at the

[3] Rearrange $\frac{\lambda_i}{(1-\lambda_i)} + \frac{k}{N} \geq 1$

choice θ_C that maximizes the expected utility is congruent with the literature on random utility. In the context of absolutely continuous random variables, a unique mode implies that the density function f_i achieves its global maximum at a single point without other global maxima on $[0,1]$. This assumption is justified by considering the shape of the expected value of the utility function in (4), which has a unique maximum and is monotonic on both sides of the maximum. The density function should mirror this shape, as the utility function represents ordinality over possible choices. Having regions of increased probability around local maximums would imply inconsistency between the deterministic preference ordering and choice probabilities in the stochastic version of the problem.

Model 0 : When bad apples are already present

In a system where only a subset of Congress gets reelected, denoted by γ, the Electorate seeks to maximize their portion θ_E. They aim to elect individuals likely to allocate close to the point of fairness. As the Electorate lacks information on individuals' complete choice probability distributions, they cannot look for agents who have a larger probability of being on the favorable side of the point of fairness ($P(\theta_C < \frac{k}{N}) = \int_0^{\frac{k}{N}} f_i(x)dx$) than others. Therefore, they base election choices on modes. Consequently, individuals with the lowest values for λ_i (modes) are elected. Suppose $k-\gamma$ individuals with $\lambda_i = 0$ are elected for the next term to Congress. However, since the Electorate has no additional information besides λ_i, their choice distributions have varying values of variances σ_i^2. Additionally, among the γ individuals not up for election, there are "bad apples" with diverse λ_i and σ_i^2 values.

The allocation decision of Congress can be modeled in two ways:

- individual choices (X_i) are made simultaneously, and $\sum_{i=1}^{k} \frac{X_i}{k}$ is the final allocation θ_C^*
- a mixture distribution $F(x) = \sum_{i=1}^{k} w_i f_i(x)$ is sampled where w_i are weights of members

Results will presumably not differ drastically, so in what follows we opt for the former.

In the above setup, the complete generality of the distributions poses difficulties in deriving meaningful relationships analytically between means, variances and being in some sense far away from the point of fairness, even though the above would be enough to confirm our later conclusions through simulations with very general distributional examples. Therefore, to simplify without fundamentally altering the problem, we introduce the following modification:

Assumption 2' (Choice probability distributions). *X_i are independent across individuals, normally distributed, with a unique mean (equivalently, mode) at $\theta_C \in [0,1]$ that maximizes the expected utilities of the individuals.*

By assuming a symmetric distribution, we imply equal receptiveness to selfish and altruistic influences. While this could be generalized, it complicates tractability without enhancing the discussion. A more significant challenge is the unbounded range of the normal distribution, whereas the choice set is limited to $[0,1]$. We address this by restricting the mean to $[0,1]$, and explaining the range by recognizing that more often than not, there are soft budgetary limits in Congress' allocation decisions. This constraint, coupled with the slim tails of the normal distribution that make large deviations from the centre of the distribution unlikely, makes it a reasonable assumption for analytical tractability.

Proposition 1 (The effect of increasingly immoral members). *All else being fixed in the composition of Congress, when the γ members already in Congress are, on average, more selfish, the probability of the mean of allocation choices θ_C^* being at distance d from the point of fairness ($\theta_C^* - \frac{k}{N} > d > 0$) grows, and the expected value of θ_E^* falls.*

Assumption 3 (Preference for certainty). *Between two final allocations θ_C' and θ_C'' for which $\mathbb{E}[\theta_C'] = \mathbb{E}[\theta_C'']$, the electorate prefers the one with smaller variance.*

The above assumption is reasonable when considering the classical literature of risk and uncertainty: with the same expected payoff, a typical individual (risk aversion) prefers the lottery with smaller variance. In this context, if one allocation is a mean-preserving spread of the other, the Electorate always prefers the one with the smaller variance.

Proposition 2 (The effect of easily swayed members). *All else being fixed in the composition* *of Congress, the Electorate is always better off when the variances of the choice probabilities of the members are smaller.*

These propositions demonstrate how different attributes of the underlying individual choice probabilities influence Congress's overall choice, and therefore impact the welfare of the Electorate. A proof for these propositions can be found in Appendix A.

The following model expands upon the significance of larger variance in choice probabilities. It also gives an explanation for findings in the cited literature above, where individuals showed a tendency to modify their initial allocation choices after being exposed to others' choices.

Model 1 : Convergence of preferences

Because everyone in Society knows the θ_C choices that maximize each individual's deterministic utility, everyone in Congress is aware of other members' selfishness. There are $t = 0, 1$ periods in Congress. Initially, an election at $t = 0$ determines Congress's composition and the allocation decision θ_C^0 (random variable). At $t = 1$ members update their parameter λ_i based on the average selfishness of Congress and their personal susceptibility, denoted by σ_i:

$$\lambda_i^1 = \frac{\lambda_i^0 + \sigma_i \lambda^0}{\sigma_i + 1} \tag{7}$$

where $\lambda^0 = \frac{1}{k} \sum_{j=1} \lambda_j^0$. This updating mechanism involves a weighted average where individuals with higher susceptibility (larger σ_i^2) are more influenced by the average selfishness of Congress. At $t = 1$, Congress makes a new allocation decision θ_C^1 after adjustments in individual selfishness. Individual variances are constant throughout the two time periods.

Let there be γ selfish members in Congress, with identical $\lambda_0^{S,i} = \lambda_0^S$ at $t = 0$. Let there be $k - \gamma$ altruistic members in Congress, with identical $\lambda_0^{A,i} = \lambda_0^A = 0$. Using the updating rule in (7) along with these assumptions, we can prove the following proposition:

Proposition 3 (The effect of easily swayed moral members and determined selfish members)**.** *All else being fixed, if individuals update their selfishness parameters λ_i in accordance with* *their level of susceptibility σ_i, then the less susceptible the immoral members of Congress are, or the more susceptible the moral members are, the worse off the Electorate is.*

The above can be interpreted in the following way: λ_0 in the $t = 0$ time period is a weighted average of the selfishness parameters. Had we not introduced updating throughout time, this weighted average would dictate the mean of the normal distribution of θ_C^*. However, with the variance-weighted updating scheme, we essentially allow the moral and immoral group to compensate their proportion in Congress ($k - \gamma$ and γ respectively) by their determination to get the allocation they want (i.e. by the size of their individual variances). As an example, say that σ_j^S are very small and that σ_i^A are very large. Then $\sum_{j=\gamma+1}^{k} \frac{\frac{1}{k}\sigma_j^S + \frac{1}{\gamma}}{1+\sigma_j^S} \simeq \gamma \cdot \frac{\frac{1}{\gamma}}{1} = 1$, $\frac{1}{k}\sum_{i=1}^{k-\gamma} \frac{\sigma_i^A}{1+\sigma_i^A} \simeq \frac{1}{k} \cdot (k - \gamma)$, and so $\left[\frac{1}{k}\sum_{i=1}^{k-\gamma} \frac{\sigma_i^A}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\frac{1}{k}\sigma_j^S + \frac{1}{\gamma}}{1+\sigma_j^S}\right] > 1$, which means that the selfish portion of Congress is able to overcompensate their proportion $\frac{\gamma}{k}$, and achieve $\lambda_1 > \lambda_0$. Similar calculations show that for small σ_i^A (determined altruistic members) $\lambda_1 < \lambda_0$. A proof of Proposition 3 can be found in Appendix A.

Model 2 : How bad apples are created

Our previous models have assumed the presence of bad apples in Congress. How bad apples end up in Congress is yet to be discussed. We could argue that the Electorate is fooled with a positive probability p about the selfishness of a candidate; the candidate gets elected under the assumption that $\lambda_i = 0$ when in reality $\lambda_i > 0$. This setup gives predictable results and is a trivial case. Much more interesting is if we assume that the voters are never fooled by the candidates.

Assume $t = 0, 1$ and that all members of Congress are altruistic ($\lambda_i = 0$) at $t = 0$. Assume further that all individuals but one use an updating formula as in (7) - clearly when all members of Congress are altruistic, there is no actual updating of the weights from period 0 to 1. Suppose there is one member in Congress who does not know the degree of selfishness of others and therefore updates his selfishness parameter based on the previous period's realization, θ_C^0, not the average selfishness parameter of other members. Using compound distributions, we can see why this is an issue for the Electorate:

Proposition 4 (The benefit of knowing all λ_j)**.** *If members of Congress are unaware of the degree of selfishness (λ_i) of each other and therefore update their selfishness parameters based on the random realization θ_C^0, the variance of the allocation choice θ_C^1 at $t = 1$ increases.*

Proposition 4 is trying to highlight the following difference: when the choice probabilities of members are "anchored" in the sense that if they share the same λ_i, there is no updating of individual selfishness weights and so the distribution of the allocation choice is stable throughout t. On the other hand, if the degree of selfishness is not common knowledge, the best information members have is the actual realization of the $t = 0$ allocation choice, θ_C^0, which is random and can potentially have a realization far away from the point of fairness $\frac{k}{N}$ even when all members are completely altruistic. The difference now lies in the fact that this random fluctuation can perturb a member's evaluation of the moral standing of Congress significantly, and so the variance of the allocation choice at $t = 1$ grows significantly. Based on Assumption 3, this makes the Electorate worse off than if the moral standing of Congress was clear and obvious to everyone.

Proof of Proposition 4 The sample mean of independently distributed normal random variables is normally distributed,

$$\theta_C^0 \sim N\left(\frac{k}{N}, \sum_{j=1}^{k} \frac{\sigma_j^2}{k^2}\right)$$

Suppose there exists a member i who is not knowledgeable about λ_j when $j \neq i$. Therefore, this member chooses θ_C^0 (and consequently, λ_i) to maximize his deterministic utility at $t = 1$: $X_i^1 \sim N\left(\theta_C^0, \sigma_i^2\right)$. This is a compound random variable of normally distributed variables, so:

$$X_i^1 \sim N\left(\frac{k}{N}, \sigma_i^2 + \sum_{j=1}^{k} \frac{\sigma_j^2}{k^2}\right)$$

From this, supposing that $X_j^1 \sim N\left(\frac{k}{N}, \sigma_j^2\right)$ as in period 0 for $j \neq i$,

$$\theta_C^1 \sim N\left(\frac{k}{N}, \sum_{j\neq 1}^{k} \frac{\sigma_j^2}{k^2} + \frac{\sigma_i^2 + \sum_{j=1}^{k} \frac{\sigma_j^2}{k^2}}{k^2}\right).$$

This is a normal distribution around $\frac{k}{N}$ with an increase in variance of $\sum_{j=1}^{k} \frac{\sigma_j^2}{k^4}$ compared to what it would be if member i could update their weight based on the average selfishness parameter of Congress. □

4 Results and possible extensions

Several extensions can be incorporated into the framework outlined above. Introducing a correlation between the degree of selfishness and the variance of individual choice probabilities can refine the results of Model 1, aligning it with empirical findings that suggest a convergence to the most selfish individuals. If the least susceptible people in Congress are coincidentally the most selfish ones, a convergence towards selfishness is more probable than a convergence towards altruism. The current model does not account for reelection incentives, which can counteract some of this observed downward convergence. Given the proven effects of individual variance on the Electorate's welfare, future models could explore how the Electorate learns an individual's susceptibility over time through repeated allocation choices, integrating this learning process into reelection decisions.

The models - built on the random choice framework - allowed us to go further in the analysis of group decision making than with a deterministic framework. Specifically, the models in this paper not only showed the effects of selfish or immoral individuals, referred to as "bad apples," within a team of decision-makers but also provided insights into why the Electorate may prefer decision-makers with strong adherence to their allocation preferences. High variability in a Congress member's choices reduces the welfare of the Electorate by increasing the variance of final allocation distributions.

Moreover, the models offer a framework for examining how individual allocation preferences converge within a team over time. By allowing an individual's susceptibility to dictate the speed of convergence in their preferences, these models are capable of adhering to diverse group dynamics, including scenarios where more selfish individuals are determined to adhere strongly to their allocation preferences. This allows them to exert a disproportionately large influence on final allocation choices, provided that other team members are sufficiently susceptible.

Finally, these models gave an explanation for how even when all individuals are altruistic in a group it can still be the case that there are larger swings towards a selfish final allocation. Introducing lack of certain information about other team member's degree of selfishness resulted in larger final allocation variances as the decision rounds progress, making the Electorate worse off than if the average selfishness of Congress was publicly known.

5 Conclusion

This work explored the dynamics of allocation choices within a team dictator game to model the decision-making process of representative bodies. Surveying the empirical research on team dictator games revealed varied outcomes, but with a general tendency toward convergence in allocation choices and a susceptibility to selfish individuals within a team. The models presented in this paper are consistent with the varying outcomes of laboratory experiments, where predictions of observed group behavior depend on each member's degree of selfishness and susceptibility to the opinions of others.

This paper presented the relationship between an elected group, Congress, and the Electorate as analogous to that between a small group of dictators and a large group of receivers, where the dictators collectively decide how to allocate a unit of good between themselves and the Electorate. Within the group of dictators, each member has their own degree of selfishness. The introduction of random individual choices made dictators susceptible to the choices of others and allowed for the modeling of negotiation processes, the swaying of individual choice, and the convergence of individual morality toward the group's average morality, with the rate of convergence determined by each member's susceptibility to surrounding preferences. The models incorporated real-world considerations, such as the presence of "bad apples" - individuals with highly selfish tendencies. The analysis revealed how the composition of Congress influences overall allocation decisions, with the presence of more selfish members leading to a higher probability of deviating from the socially accepted allocation choice i.e. the point of fairness. Results showed that higher degree of susceptibility among Congress members to other preferences negatively impacts the welfare of the Electorate.

A Proofs

Choice solution of a member of Congress:

First, let $f(\theta) = a_i \frac{\theta_C}{k} - \frac{b_i}{2k}\left(\theta_C - \frac{k}{N}\right)^2$ while $a_i \geq 0, b_i > 0$. Then using the chain rule in the second additive component of f and setting f' to zero, we get

$$f' = \frac{a_i}{k} - \frac{b_i}{2k}(2\theta - \frac{2k}{N}) = \frac{a_i}{k} - b_i(\frac{\theta}{k} - \frac{1}{N}) = 0.$$

Rearranging gives $\theta = \frac{a_i}{b_i} + \frac{k}{N}$. Since the value of θ is restricted to $[0, 1]$, the solution can be expressed as $\theta^* = \min\{1, \quad \frac{a_i}{b_i} + \frac{k}{N}\}$. (To convince ourselves that this is maximizing, notice that $f'' = -\frac{b_i}{k}$ where both $b_i, k > 0$.) Now, assume $b_i = 0$. In this case, f grows linearly in θ and the trivial solution is $\theta_C^* = 1$.

Proof of Propositions 1 and 2:

We have $k - \gamma$ independent random variables $X_i \sim N(\frac{k}{N}, \sigma_i^2)$, and γ random variables $X_j \sim N(\frac{\lambda_j}{(1-\lambda_j)} + \frac{k}{N}, \sigma_j^2)$ (assuming $1 - \frac{1}{2-\frac{k}{N}} > \lambda_i \geq 0$). Then

$$\begin{aligned}
\sum_{i=1}^{k} X_i = \sum_{j=1}^{\gamma} X_j + \sum_{i=\gamma+1}^{k} X_i \qquad & \sim N\left(k \cdot \frac{k}{N} + \sum_{j=1}^{\gamma} \frac{\lambda_j}{(1-\lambda_j)}, \sum_{j=1}^{k} \sigma_j^2\right) \\
\theta_C^* = \frac{1}{k}\sum_{i=1}^{k} X_i \qquad & \sim N\left(\frac{k}{N} + \frac{1}{k} \cdot \sum_{j=1}^{\gamma} \frac{\lambda_j}{(1-\lambda_j)}, \sum_{j=1}^{k} \frac{\sigma_j^2}{k^2}\right)
\end{aligned}$$

Proposition 1: The mean selfishness of the γ members is $\frac{1}{\gamma} \cdot \sum_{j=1}^{\gamma} \frac{\lambda_j}{(1-\lambda_j)} \propto \frac{1}{k} \sum_{j=1}^{\gamma} \frac{\lambda_j}{(1-\lambda_j)} = S$

which sets the mean of θ_C^* to $\frac{k}{N}+S$. Let $\sum_{j=1}^{k}\frac{\sigma_j^2}{k^2}=\sigma^2$. Then

$$\mathbb{E}\left[\theta_E^*\right]=\mathbb{E}\left[1-\theta_C^*\right]=1-\left(\frac{k}{N}+\frac{1}{k}\cdot\sum_{j=1}^{\gamma}\frac{\lambda_j}{(1-\lambda_j)}\right)=1-\frac{k}{N}-S\searrow \text{ as } S\nearrow.$$

$$P(\theta_C^*-\frac{k}{N}>d)=\int_{\frac{k}{N}+d}^{\infty}f_{\theta_C^*}(x)dx=1-\int_{-\infty}^{\frac{k}{N}+d}f_{\theta_C^*}(x)dx=1-0.5\left[1+\text{erf}\left(\frac{\frac{k}{N}+d-(\frac{k}{N}+S)}{\sqrt{2}\sigma}\right)\right]=$$

$$=1-0.5\left[1+\text{erf}\left(\frac{d-S}{\sqrt{2}\sigma}\right)\right]=0.5-\frac{1}{\pi}\int_{0}^{\frac{d-S}{\sqrt{2}\sigma}}e^{-t^2}dt\nearrow \text{ as } S\nearrow.$$

where the last conclusion comes from e^{-t^2} being strictly positive on the positive interval, therefore decreasing the length of $([0,\frac{d-S}{\sqrt{2}\sigma}])$ decreases the the integral.

Proposition 2: This is a direct consequence of Assumption 3: for each individual member's choice X_i, increasing σ_i^2 increases the variance of the random variable θ_C^*, $\sum_{i=i}^{k}\frac{\sigma_i^2}{k^2}$, which by assumption means the Electorate is worse off due to more uncertainty. □

PROOF OF PROPOSITION 3: The mean selfishness parameter at $t=0$ in Congress is:

$$\lambda_0=\frac{\gamma}{k}\lambda_0^S+\frac{k-\gamma}{k}\lambda_0^A=\frac{\gamma}{k}\lambda_0^S.$$

The updated selfishness parameters at $t=1$ are:

$$\lambda_1^{A,i}=\frac{\lambda_0^A+\sigma_i^A\lambda_0}{1+\sigma_i^A},\quad \lambda_1^{S,i}=\frac{\lambda_0^S+\sigma_i^S\lambda_0}{1+\sigma_i^S}$$

and the mean selfishness parameter at $t = 1$ in Congress is:

$$
\begin{aligned}
\lambda_1 &= \frac{1}{k}\left[\sum_{i=1}^{k-\gamma} \frac{\lambda_0^A + \sigma_i^A \lambda_0}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\lambda_0^S + \sigma_j^S \lambda_0}{1+\sigma_j^S}\right] \\
&= \frac{1}{k}\left[\sum_{i=1}^{k-\gamma} \frac{\lambda_0^A}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\lambda_0^S}{1+\sigma_j^S} + \lambda_0 \left(\sum_{i=1}^{k-\gamma} \frac{\sigma_i^A}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\sigma_j^S}{1+\sigma_j^S}\right)\right] \\
&= \frac{1}{k}\left[\lambda_0^S \left(\sum_{j=\gamma+1}^{k} \frac{1}{1+\sigma_j^S}\right) + \lambda_0 \left(\sum_{i=1}^{k-\gamma} \frac{\sigma_i^A}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\sigma_j^S}{1+\sigma_j^S}\right)\right] \\
&= \frac{1}{k}\left[\lambda_0^S \left(\sum_{j=\gamma+1}^{k} \frac{1}{1+\sigma_j^S}\right) + \frac{\gamma}{k}\lambda_0^S \left(\sum_{i=1}^{k-\gamma} \frac{\sigma_i^A}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\sigma_j^S}{1+\sigma_j^S}\right)\right] \\
&= \frac{\gamma}{k}\lambda_0^S \left[\frac{1}{k}\sum_{i=1}^{k-\gamma} \frac{\sigma_i^A}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\frac{1}{k}\sigma_j^S + \frac{1}{\gamma}}{1+\sigma_j^S}\right] \\
&= \lambda_0 \left[\frac{1}{k}\sum_{i=1}^{k-\gamma} \frac{\sigma_i^A}{1+\sigma_i^A} + \sum_{j=\gamma+1}^{k} \frac{\frac{1}{k}\sigma_j^S + \frac{1}{\gamma}}{1+\sigma_j^S}\right]
\end{aligned}
$$

With all else fixed, if σ_j^S decreases for any of the selfish members, $\frac{\frac{1}{k}\sigma_j^S+1}{1+\sigma_j^S}$ increases because $\frac{d}{d\sigma_j^S}\frac{\frac{1}{k}\sigma_j^S+1}{1+\sigma_j^S} = \frac{\frac{1}{k}-\frac{1}{\gamma}}{(\sigma_j^S+1)^2} < 0$. Furthermore, $\frac{d}{d\sigma_j^A}\frac{\frac{1}{k}\sigma_j^A}{1+\sigma_j^A} = \frac{\frac{1}{k}}{(\sigma_j^A+1)^2} > 0$. These show the relationship between the size of λ_1 and the variances of the altruistic and selfless groups. Since λ_1 determines how far the mean of the normally distributed θ_C^* is from the point of fairness, and since the variances are kept fixed between $t = 0, 1$, the Electorate clearly prefers a smaller λ_1, equivalently, $\mathbb{E}[\theta_C^*]$ closer to $\frac{k}{N}$. $\square$

References

Agranov, M., & Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, *125*(1), 40–68.

Barberá, S., & Pattanaik, P. K. (1986). Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, 707–715.

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy*, *82*(6), 1063–1093.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, *8*(1), 41–55.

Bolton, G. E., & Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, *91*(1), 166–193.

Cason, T. N., & Mui, V.-L. (1997). A laboratory study of group polarisation in the team dictator game. *The Economic Journal*, *107*(444), 1465–1483.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, *100*(2), 193–201.

Dillenberger, D., & Sadowski, P. (2012). Ashamed to be selfish. *Theoretical Economics*, *7*(1), 99–124.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*, 583–610.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—intentions matter. *Games and Economic Behavior*, *62*(1), 287–303.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, *18*(1), 52–72.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

Franzen, A., & Pointner, S. (2014). Giving according to preferences: Decision-making in the group dictator game. *Soziale Welt*, *65*(2), 139–152.

Goeree, J. K., McConnell, M. A., Mitchell, T., Tromp, T., & Yariv, L. (2010). The 1/d law of giving. *American Economic Journal: Microeconomics*, *2*(1), 183–203.

Gul, F., & Pesendorfer, W. (2006). Random expected utility. *Econometrica*, *74*(1), 121–146.

Ito, T., Suzuki, A., Takemoto, T., Ogawa, K., & Takahashi, H. (2016). Contagion of self-interested behavior: Evidence from group dictator game experiments. *German Economic Review*, *17*(4), 425–437.

Kerr, N. L., Rumble, A. C., Park, E. S., Ouwerkerk, J. W., Parks, C. D., Gallucci, M., & Van Lange, P. A. (2009). "how many bad apples does it take to spoil the whole barrel?": Social exclusion and toleration for bad apples. *Journal of Experimental Social Psychology*, *45*(4), 603–613.

Leider, S., Möbius, M. M., Rosenblat, T., & Do, Q.-A. (2009). Directed altruism and enforced reciprocity in social networks. *The Quarterly Journal of Economics*, *124*(4), 1815–1851.

Luhan, W. J., Kocher, M. G., & Sutter, M. (2009). Group polarization in the team dictator game reconsidered. *Experimental Economics*, *12*, 26–41.

Malmberg, H. (2013). *Random choice over a continuous set of options* [Doctoral dissertation, Department of Mathematics, Stockholm University].

Malmberg, H., & Hössjer, O. (2018). Continuous approximations of discrete choice models using point process theory. *Stochastic Processes and Applications: SPAS2017, Västerås and Stockholm, Sweden, October 4-6, 2017*, 413–435.

McFadden, D., et al. (1974). Frontiers in econometrics, chapter conditional logit analysis of qualitative choice behavior.

McFadden, D. L. (2005). Revealed stochastic preference: A synthesis. *Economic Theory*, *26*(2), 245–264.

Panchanathan, K., Frankenhuis, W. E., & Silk, J. B. (2013). The bystander effect in an n-person dictator game. *Organizational Behavior and Human Decision Processes*, *120*(2), 285–297.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 1281–1302.