# Social Influence in Legal Deliberations

**Chaim Fershtman**[†] **Uzi Segal**

**September 12, 2021**

**Abstract**

**Juries, appellate courts, parole boards are all institutes that need to make collective decisions. What characterizes these institutes is that they are typically engage in deliberations prior to decision making. Beyond information exchange,monstrate**

**t the voting outcome even when**
**then demonstrate the ability of**
**procedure in order to increase**
**will be selected.**

**ence, deliberation.**

[*] **We thank Oren Bar-Gill, Anat Horovitz, and Ariel Porat for valuable comments and**

# 1 Introduction

**Deliberation characterizes most of collective decision making. Examples can be a jury, parole boards, congressional committees, multi-judge appellate courts, or in general, any committee that needs to make a collective decision.[1] Di□erent institutions may have di□erent deliberation protocols. For example, deliberations among the Justices in the US Supreme Court proceed by order of seniority.[2] During such deliberations individual decision makers exchange information, argue, exchange ideas, and try to persuade and convince each other regarding the "right" decision. Final decisions, which are often made by a vote, are the outcomes of the collective interactions during the deliberation process.**

In a previous paper (Fershtman and Segal (2018), hereafter FS) we modeled social influence by introducing a setup in which each individual is characterized by two sets of preferences: unobservable core preferences and observable behavioral preferences, where actual choice is determined by the latter. Each person has an individual social influence function that determines the way this individual is a□ected by the opinions of others. Formally,

**replacement of one of its members with another juror who in her core preferences prefers alternative $A$. We show that it is possible that such a switch will result in a shift of vote by the jury from $A$ to $B$. This may happen if the replaced juror has strong preferences for $A$ while the new juror's preferences for $A$ are much milder and therefore she will be much less e□ective in the social influence process. In a similar way we show that if a committee is expected to vote for one of the alternatives then adding a member who in her core preferences prefers the same alternative may induce the committee to change its opinion. Finally, we show that as a result of social influence, deliberation may result in a violation of the unanimity property. That is, even when all committee members prefer in the196(i)-0.765963-17.334Td [(s)-0.19416.5750Td [(e)-**

# 2 The Model

## 2.1 Preliminaries

**A group of decision mes**

no alternative receives unanimous support, then typically there is a variety of options how to proceed, for example, instructing the juries to continue deliberation or declaring a mistrial (see e.g. LaFave et al. §

of interaction is irrelevant to jury, judges, or parole-boards deliberations, as both are forbidden from using any private information. In contrast, the social influence we discuss is about a□ecting preferences, not information.

When juries need to determine conviction, and for this choice they need to weigh di□erent types of evidence, it is not clear that reac

own core preferences and that the dependence on other people's observed behavior is of acceptance and not of rejection. The requirements $g_1 \leq 1$ and $g_2 \leq 1$ mean that the change in the behavioral parameter cannot be larger than the change in the relevant parameters, reflecting the fact that the other parameters which did not change mitigate the influence of changes of the parameter that did change. Finally, $g_{12} < 0$ suggests that the sensitivity of a person's behavior to an increase in his core preferences is higher when these preferences are moving in an opposite direction to his observed environment as such a change is more indicative to him than when his core preferences move up together with the observed preferences of everyone else.

Interactions between individuals open the door to possible manipulations and strategic behavior. In the present context, there are two possible types of such behavior. Individuals may misrepresent their views, knowing that other members are influenced by their discourse and arguments, and those who control the procedures of deliberation may manipulate it in order to influence its outcome. In this paper we want to focus attention on the second type of strategic behavior. We assume that the organizers of the deliberation procedures understand the pattern of social influence and may manipulate the deliberation procedure in order to affect its outcome. We assume, however, that committee members themselves express their true opinions without any strategic motives. What we have in mind is a parole board or judges who make periodic decisions on various issues and its members cannot express different opinions at different meetings. On the other hand, the chair of the committee has the power to determine the procedure of deliberation, for example, its order, and it may change from one meeting to another.

## 2.3 Networks of Influence

When there is a jury in which there is a free discussion without any specific protocol of deliberation then we are looking for the equilibrium profile of behavioral preferences such that the behavioral preferences of each individual is derived from her core preferences and the behavioral preferences of other jury

# 3 Social influence in Jury Deliberation

**Jury deliberation is typically done without specific protocol. There is a discussion without any specific order of speaking and all jury members may participate in the deliberation, which is then followed by voting. In this section we discuss the relationship between the profile of jury members and the final vote, and define several intuitive properties that the deliberation process may satisfy. These properties are satisfied when jury members vote according to their core preferences. But whether these properties are satisfied at the presence of social influence depends on the individual social influence functions as well as on the profile of their core preferences.**

**We start with a simple property of unanimous acceptance which requires that if there is a unanimous support for an alternative prior to the deliberation, then it will be chosen by the jury after the deliberation as well. Note that this is a weak notion of unanimous acceptance. A stronger version would imply that if prior to the deliberation all jury members prefer one alternative to another, then after deliberation they will still** $n\ n_{|}\ o\ s$ **vote for the first. The justification for the stronger version is that if no jury member supports a certain alternative, then no one will be able to convince others to vote for it.**

**Property 1 (Unanimous Acceptance): If all jury members prefer one alternative (e.g. for all** $i$**,** $\alpha$

**The above claim provides a valuable information regarding the relationship between the types of jury members and their final vote. Its main message is that it is not enough to focus on the ordinal preference (i.e., which alternative the juror prefers) as the cardinal preferences (the intensity of the ordinal preferences) play an important role in determining the social influence and therefore the final vote.**

on two aspects of the deliberation procedure: ($i$) the e□ect of the order of

six orders ($i$): 1-2-3, ($ii$): 1-3-2, ($iii$): 2-1-3, ($iv$): 2-3-1, ($v$): 3-1-2, and ($vi$): 3-2-1.

**Claim 4** Let $\alpha_1 < \alpha_2 < \alpha_3$. If decision is made by the unanimity rule, then option $B$, which has an advantage in the second attribute, is most likely to be selected if the order of deliberation is 3-1-2, regardless of the value of $\theta$. Under majority rule, it is most likely to be elected if the order is 3-2-1.

This claim shows that the designer of the protocol, who has his own favorite choice, may benefits from manipulating the order in which deliberation takes place. However, the optimal order depends on the committee's decision rule. There are different optimal orders under majority and unanimity rules.

**Remark:** It is clear from Table 1 in the proof of Claim 4 that the order of the behavioral parameters $\beta_1, \beta_2, \beta_3$ does not have to be the same as the order of the core parameters $\alpha_1, \alpha_2, \alpha_3$. For example, if $g(\alpha_3, \alpha_1) < \alpha_2$ and $\theta$ is sufficiently close to 0, then $\beta_2^{(ii)} > \beta_3^{(ii)}$ even though $\alpha_2 < \alpha_3$ (see Table 1).

## 4.2 The Effect of No Participation in the Deliberation

In many committees members do not have to participate in the deliberation. They can vote without explaining their opinion or they may even send their vote by mail without listening to the opinions of other committee members. In order to demonstrate the effect of such procedures we consider a decision making by a parole board, where (one of) the relevant factors is the safety of the community. Suppose for simplicity that the board specifies a benchmark $\gamma$ of a critical risk level such that any prisoner posing a higher risk to society will not be granted a parole. However, each board member has her own opinion regarding the risk associated with each prisoner. Although all members have the same information and are subject to the same guidelines, they may differ in the way they apply these guidelines and information to specific cases, following their different experiences or different backgrounds,

captured by a single parameter $\alpha$ (which is equivalent in our terminology to the core preferences). Similarly, following the deliberation in the parole board each member may adjust her opinion to $\beta$ (which is equivalent to the behavioral preferences). Thus, members with $\beta \square \gamma$ will vote in favor of a parole and those with $\beta > \gamma$ will vote against it.

We assume boards of three members and consider two possible decision rules. The first is a majority rule in which a parole decision requires the support of at least two members of the board. The second is a unanimity

**are possible.**

**right choice. This is true in juries, parole boards and even in most families. Our paper focuses on the e□ect of deliberation as a mechanism that changes preferences and opinions. This important aspect of deliberation implies that the deliberation process is not just an exchange of information (or manipu-**

$\alpha$. Likewise, if $g(\alpha, \alpha, \ldots, \alpha) > \alpha$, then $\alpha < g(\alpha - \varepsilon, \alpha, \ldots, \alpha)$, a contradiction. □

**Proof of Claim 3:**

**Monotonicity:** Consider the system $\beta_i = g^i(\alpha_i, \sum_{j \neq i} \beta_j/(n-1))$, $i = 1, \ldots, n$. Take the total differential to obtain for $i = 1, \ldots, n$

$$g_1^i \left( \alpha_i, \sum_{j \neq i} \frac{\beta_j}{n-1} \right) = \frac{d\beta_i}{d\alpha_i} - \frac{1}{n-1} \sum_{j \neq i} g_2^i \left( \alpha_i, \sum_{j \neq i} \frac{\beta_j}{n-1} \right) \frac{d\beta_j}{d\alpha_i} \qquad (1)$$

Let the matrix B be given by $b_{i,i} = 1$, and $b_{i,j} = -\frac{1}{n-1} g_2^i \left( \alpha_i, \sum_{j \neq i} \frac{\beta_j}{n-1} \right)$ whenever $i = j$. Let $C_j$ be obtained from B by replacing column $j$ of B with $\left( 0, \ldots, 0, g_1^i \left( \alpha_i, \sum_{j \neq i} \frac{\beta_j}{n-1} \right), 0, \ldots, 0 \right)^T$. The matrices B, $C_1, \ldots, C_n$ all satisfy the conditions of theorem 4.D.1 in Takayama (1985, p. 392), and moreover, for $x^T = (1, \ldots, 1)$ and $A = B, C_1^T, \ldots, C_n^T$, $Ax \neq 0$ (recall that $0 < g_2 < 1$). By the above theorem, $\det(B), \det(C_1), \ldots, \det(C_n) > 0$. It thus follows from the system of linear equations (1) that for all $i, j$, $\frac{d\beta_j}{d\alpha_i} > 0$. All committee members are now more inclined to choose alternative $B$, and as it was preferred to $A$ before the shift, it is certainly preferred after.

**Unanimous Acceptance:** Suppose that all members have the same social influence function $g(\alpha, \beta)$ such that $\beta$ is the average preferences of everyone else. If all agents have the same core preferences $\alpha$ and the social preference function is SR (i.e., $g(\alpha, \alpha) > \alpha$), then the equilibrium occurs at $\beta > \alpha$ (see Claim 6 in FS). Let $\alpha'$ and $\beta'$ be such that $\beta' = g(\alpha', \beta')$. If $\alpha' < \gamma < \beta'$ then by their core preferences all agents prefer $A$ to $B$ (since $\alpha' < \gamma$), but by their behavioral preferences they would vote for $B$ since $\gamma < \beta'$.[9]

**Consistency:** Suppose that all members have the same core preferences $\alpha > \gamma$, but as their preferences are UR (that is, $g(\alpha, \alpha) < \alpha$), their common

**behavioral preferences $\beta$ are just above $\gamma$ and alternative $B$ is selected. Add a new committee member whose core preferences are just above $\gamma$ (but sufficiently below $\alpha$) and his preferences may push the behavioral preferences of all other agents below $\gamma$.** □

**Proof of claim 4**

**($vi$) 3-2-1:** $\beta_3^{(vi)} > \beta_2^{(vi)} > \beta_1^{(vi)}$**, hence** $M^{(vi)} = \beta_1^{(vi)}$**. Obviously** $M^{(vi)} \square \beta_1^{(v)} < \beta_3^{(v)}$ **and by definition,** $M^{(vi)} \square \beta_1^{(vi)}$**. We show next that for all** $\alpha_2 \in [\alpha_1, \alpha_3]$**,** $\beta_1^{(vi)} \square \beta_2^{(v)}$

**Suppose that** $\beta_2 > \alpha_2$ **but** $\beta_2^1 \sqsupset \beta_2$**. Since by FS** $\beta_1 < \beta_3$ **(see end of subsection 2.2),**

$$\beta_2 = g(\alpha_2, \tfrac{1}{2}[\beta_1 + \beta_3]) \sqsupset \beta_2^1 = g(\alpha_2, \beta^1$$

more project it is enough to show that $\beta_2 > \beta_2^3$. Since by the aforementioned claim, $\alpha_2 > \beta_2^3$, this is clearly the case when $\beta_2 \square \alpha_2$. We therefore prove the impossibility of $\alpha_2 > \beta_2^3 > \beta_2$. Otherwise,

$$\beta_2^3 = g(\alpha_2, \beta_1^3) \square \beta_2 = g(\alpha_2, \tfrac{1}{2}[\beta_1 + \beta_3]) = \quad \beta_1^3 > \beta_1$$

Since $g_2 < 1$, we get

$$\left.\begin{array}{r} \beta_2^3 - \beta_2 < \beta_1^3 - \frac{1}{2}[\beta_1 + \beta_3] \\ \beta_1^3 - \beta_1 < \beta_2^3 - \frac{1}{2}[\beta_2 + \beta_3] \end{array}\right\} =$$
$$2\beta_3 < \beta_1 + \beta_2$$

A c02Td [(.)-62=□7.984.5553910Td [(□)-0.575883]TJ R2d [(fi)0.434p [(fi)0.434975]TJ R257.9701

# References

[1] Aronson, E., T.D. Wilson, and R.M. Akert, 2010. Social Psychology, 7th ed. Upper Saddle River: Prentice Hall.

**[11] Fershtman, C. and U. Segal, 2018. "Preferences and social influence,"**

[21] Moldovanu, B. and X. Shi, 2013. “Specialization and partisanship in committee search,” Theoretical Economics 8:751–774.

[22] Myers, D.G., 1975. “Discussion-induced attitude polarization,” Human Relations 28:699–714.

[23] Myers, D., 1982. “Polarizing e□ects of social interaction,” In H. Brandstatter, J. Davis, and G. Stocker-Kreichbauer (Eds.), Group Decision