Voting in large committees with disesteem payoffs: A 'state of the art' model *

Rune Midjord, Tomás Rodríguez Barraquer, and Justin Valasek†

February 10, 2017

Abstract

In this paper, we consider a committee of experts that decides whether to approve or reject a proposed innovation on behalf of society. In addition to a payoff linked to the correctness of the committee's decision, each expert receives disesteem payoffs if either he/she votes in favor of an ill-fated innovation (a type I error) or votes against an innovation that proves to be beneficial (a type II error). We find that the predictions of the model are sensitive to the assumed signal technology. The standard Condorcet framework assumes that experts' signals are i.i.d. conditional on the state of the world, implying that the state of the world is approximated with arbitrary precision by a sufficiently large number of signals. Surprisingly, with this assumption, any combination of disesteem payoffs leads to large committees accepting the innovation with too high a probability. However, if this assumption is relaxed, then depending on the relative size of the disesteem payoffs the committee may accept or reject the innovation with too high a probability.

Keywords: Committees, Information aggregation, Disesteem payoffs.
JEL Classification Codes: D71, D72

*We thank David Austen-Smith, Salvador Barberà, Matthew Ellman, Sivan Frenkel, Andrea Galeotti, Volker Hahn, Sergiu Hart, Steffen Huck, Macartan Humphreys, Nagore Iriberri, Annick Laruelle, Marcos Nakaguma, Henry Mak, Andrea Matozzi, Rebecca Morton, Peter Norman Sorensen, Jesús Vázquez, Fernando Vega-Redondo, Bauke Visser, Elyashiv Wiedman, Huseyin Yildirim and seminar participants at EUI, Public University of Navarra, the University of the Basque Country, the Center for Rationality at Hebrew University, University of Bern, IGIER-Bocconi, and the AEA Annual Meeting. Tomás Rodríguez Barraquer acknowledges funding for the research leading to these results from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n. [249159]. Furthermore, he would specially like to thank the hospitality and support of the Center for the Study of Rationality at Hebrew University during the development of this project. All errors are our own.

†University of Copenhagen, Universitat Autònoma de Barcelona, and WZB Berlin respectively. Contact: Rune Midjord, Oester Farigmagsgade 5, 1353 Copenhagen. Telefon: +4535323059. Email: rune.midjord@econ.ku.dk.

1 Introduction

The rationale for delegating a decision to a group of experts rather than an individual is clear: committees aggregate multiple sources of information and expertise, and therefore allow for more informed decisions. However, by participating in a committee, experts may face idiosyncratic payoffs tied to the correctness of their personal vote. An example is FDA committees, where committee members may be exposed to a negative payoff if they vote to approve a drug that proves to be fatal for some users, or vote against a drug that successfully treats a previously incurable illness. For instance, when Posicor, a drug to treat high blood pressure, resulted in the death of over 140 people, numerous newspaper articles (including an article that received the prestigious Pulitzer Prize) singled out individual committee members based on their vote – while the committee as a whole made the wrong decision, only committee members who personally voted for the drug were scrutinized.

In this paper we analyze committee behavior when, in addition to caring that the committee makes the right decision, each committee member faces a negative disesteem payoff if his/her individual vote is shown to differ from the appropriate choice.[1] As in the FDA example, we consider a committee that decides whether or not to adopt an innovation, in an environment where the quality of the innovation becomes evident only if it is adopted. Because of this one-sided revelation of quality, committee members are only exposed to disesteem payoffs when the innovation is accepted.

In this environment, we find that disesteem payoffs generically distort the decision away from perfect information aggregation in large committees. When the disesteem payoff for voting to reject a good innovation (type II error) is large relative to the payoff for voting to accept a bad innovation (type I error), then the predictions of the model are intuitive and the committee will vote to accept the innovation with too high a probability.

However, when the disesteem payoff for a personal type I error is large relative to the payoff for a personal type II error, the predictions of the model depend on the technology that generates the private information of the committee members. In our model, each committee member receives a private signal that indicates that the innovation is either good or bad. The Condorcet framework, which is the standard model used to analyze information aggregation in committees, assumes that each signal is i.i.d. conditional on the true state of the world. This implies that the aggregation of information held by a sufficiently large group of individuals reveals the state of the world with arbitrary precision. In contrast, we consider a model that includes the Condorcet framework as a special case, but that also allows for

[1]This payoff can be purely intrinsic (self-esteem), or as in Brennan and Pettit (2004) and Ellingsen and Johannesson (2008), esteem payoffs can reflect an agent's payoff from their general regard by other members of society (also see the discussion of the relevant psychological and classical literature in Brennan and Pettit).

an alternative signal technology. Specifically, our model also considers the case where each expert's signal is i.i.d. conditional on a *state of the art*, a random variable that equals the state of the world with high probability, but which may also be incorrect. This implies that the aggregation of information held by a large group of individuals conveys the false state of the world with a probability that is bounded away from zero.

We show that under the standard Condorcet signal technology (*state of the world*), a large committee of experts will always act rashly, accepting the innovation with too high a probability. That is, no matter how large the disesteem payoff for voting to accept a bad innovation, there is no over-caution in large committees of experts. This finding, while interesting in and of itself, is not robust: under the alternative state-of-the-art signal technology, where the collective knowledge contained in even a very large number of signals has some probability of being wrong, when the disesteem payoff for personal type I error is relatively large, a large enough committee will always reject the innovation regardless of the information held by its members.

To see the intuition behind this difference in the state of the world and state of the art models, consider the case of a large committee where the disesteem payoff for voting to accept a bad innovation is relatively large. One might expect this to give rise to over-caution under the state of the world model: if the committee accepts the innovation then personal errors are harshly punished only for those who vote to accept. However, if the committee is over-cautious, then the (large) committee practically never approves a bad innovation, which eliminates the impact of the payoff for a type I error. In contrast, this intuition fails in the state of the art model since there is always a positive probability that the state of the art is wrong, which implies that over-caution can be the unique equilibrium given a large relative payoff for a type I error.

The paper is organized as follows. Following a review of the literature, section 2 introduces the payoff structure and the process that generates each expert's opinion (signal). Section 3 characterizes the limit results of the general state-of-the-art model, and compares the state of the world and state of the art models. All proofs are relegated to the appendix. In a supplementary Appendix, available online, we present suggestive evidence that larger committees reject innovations more frequently using data on the voting patterns of FDA committees, and include an analysis of information aggregation under the state of the art view of expertise without disesteem payoffs, which is a special case of our model.[2]

[2] All the results in the absence of disesteem payoffs are analogous to those of the literature on the Condorcet jury theorem with strategic voters (see Austen-Smith and Banks (1996)), McLennan (1998) and Feddersen and Pesendorfer (1998)). For a general version of the Condorcet jury theorem, see Peleg and Zamir (2012).

Literature Review

This paper contributes to the game theoretic literature on information aggregation in committees (see Austen-Smith and Banks (1996) for an early reference and recent surveys by Gerling et al. (2005) and Li and Suen (2009)). Our paper is closely related to a subset of the committee literature that considers information aggregation when voters have a common interest in making the right decision and additional "idiosyncratic" payoffs that condition on the individuals' votes.[3]

In Visser and Swank (2007), committee members deliberate on whether to accept a project prior to voting. The members are concerned about the value of the project and their reputation for being well informed. The market, whose judgement the experts care about, does not observe the value of the project, only the decision taken by the committee. One difference with respect to our model is that in Visser and Swank the additional reputation (idiosyncratic) payoffs do not directly depend on the state (the true value of the project is never revealed). Callander (2008) analyzes idiosyncratic payoffs in elections when voters wish for the better candidate to be elected, but also to personally vote for the winner. The payoff for voting for the winner (independently of the winning candidate's quality) creates multiple symmetric equilibria, some of which have unusual properties. Huck and Konrad (2005) show that delegating a decision to a large committee whose members face a small moral cost when voting in a particular direction can function as a commitment device, since a large committee will always vote in the direction of the "moral" option. Morgan and Várdy (2012) study a model in which voters are driven by both instrumental and purely expressive idiosyncratic payoffs. That is, a voter receives some consumption utility if he/she votes in a pre-defined way (e.g. in accordance with one's norms) that is irrespective of the correct outcome and the implemented decision.

While Callander (2008) and Morgan and Várdy (2012) both demonstrate that idiosyncratic payoffs can lead to a failure of information aggregation in large committees, the mechanism we present here is quite different. In both of the above papers, idiosyncratic payoffs give agents a direct incentive to vote for, say, candidate A regardless of the state of the world; that is, information aggregation fails when idiosyncratic payoffs run counter to the common value payoff of electing the better candidate. In our analysis, however, information aggregation fails despite idiosyncratic payoffs that reinforce common value payoffs: disesteem payoffs realize only for experts who vote to approve a bad drug, or vote against approving a good drug.

[3] In another branch of the literature the committee members have no concern for the aggregate decision and care only about voting (or giving recommendations) to maximize the belief that the "market" holds about their level of competence – i.e. the precision of their private signals. See e.g. Ottaviani and Sorensen (2001) and Levy (2007).

Lastly, Li (2001) shows that committees might have an incentive to adopt a more conservative decision rule, in the sense of requiring a higher information threshold, to induce members to individually invest more in information gathering. Our results give a complementary explanation for why, even in situations where the committee decision rule is based on votes rather than quantifiable evidence, committee members have an incentive to vote conservatively.

2 The Model

An innovation is submitted for approval by a committee of n experts that operates according to a q-rule: If strictly more than a fraction q of the committee members $i \in \{1, 2, ..., n\}$ vote in favor of approval then the innovation is approved, and otherwise it is rejected. We denote the vote of committee member i by $v \in \{a, r\}$ and the decision of the committee (outcome) by $o \in \{a, r\}$, where a indicates *accept* and r indicates *reject*.[4] The payoff to each expert i depends on the decision of the committee, an underlying state of the world $\omega \in \{A, R\}$, and the expert's vote v:

$$U(v, o, \omega) = \begin{cases} 0 & \text{if } o = r \\ w & \text{if } o = a,\ \omega = A,\ v = a \\ w - k_2 & \text{if } o = a,\ \omega = A,\ v = r \\ -c & \text{if } o = a,\ \omega = R,\ v = r \\ -c - k_1 & \text{if } o = a,\ \omega = R,\ v = a \end{cases}$$

where $w, c, k_1, k_2 \geq 0$.

One interpretation of the structure of the payoffs is as follows: if the innovation is rejected, then payoffs to all agents in the committee are zero, since the status quo is preserved and no further information about the innovation's quality is generated. If the innovation is approved, then the quality of the innovation is revealed and the committee members receive a common payoff and, depending on the state of the world and their vote, an individual disesteem payoff. The common payoff is w or $-c$ depending on whether the committee has made the right decision with respect to the state of the world. The individual disesteem payoff depends on whether the individual has made a type I or a type II error in his expressed judgment. Specifically, if the committee has wrongly accepted the innovation and the committee member has voted to accept (which we will refer to as a type I error) the disesteem payoff amounts to k_1. It amounts to k_2 if the committee has correctly accepted

[4] We consider any q-rule with a fixed q, such as the majority rule used in FDA committees. This excludes decision rules such as the unanimity rule, where $q = (n-1)/n$; for an analysis of the case of unanimity and communication, see a working version of the paper, Midjord et al. (2014).

the innovation, yet the committee member has voted to reject (which we will refer to as a type II error).[5]

We denote by $p_A \equiv \Pr[\omega = A]$ the experts' common prior belief on the state of the world. In addition to the state of the world, we introduce the concept of the *state of the art*, denoted by $\tau \in \{a, r\}$. The state of the art is generated by the underlying state of the world, with some probability of error. Let e_1 denote the probability that the state of the art is wrong when it indicates that the innovation should be rejected ($e_1 = \Pr[\omega = A|\tau = r], 0 \leq e_1 < \frac{1}{2}$), and let e_2 denote the probability that the state of the art is wrong when it indicates that the innovation should be accepted ($e_2 = \Pr[\omega = R|\tau = a], 0 \leq e_2 < \frac{1}{2}$).

Neither the state of the world, nor the state of the art is directly observable to the experts. Instead, each expert receives a private signal, $s \in \{a, r\}$, that is generated by the state of the art; that is, the experts' signals are i.i.d. conditional on the state of the art. With probability $1 - \varepsilon$ the signal of expert i coincides with the state of the art ($\Pr[s = \tau] = 1 - \varepsilon$, $\varepsilon < \frac{1}{2}$), and with probability ε it differs with respect to the state of the art ($\Pr[s \neq \tau] = \varepsilon$).

We introduce the state of the art to reflect the possibility that even an infinite number of expert signals may not reveal the true state of the world – i.e. even the ideal analysis of all available data may not reveal the true quality of a proposed innovation. The standard Condorcet framework, where signals are generated directly by the state of the world, corresponds to the case of $e_1 = e_2 = 0$. For expositional clarity, we refer to the standard model ($e_1 = e_2 = 0$) as the SoW model, and the state of the art model ($e_1, e_2 > 0$) as the SoA model.

In what follows we will use $\sigma = (\sigma_a, \sigma_r)$, to denote the possibly-mixed strategy according to which member i sets $v = a$ with probability $0 \leq \sigma_a \leq 1$ after receiving signal $s = a$, and sets $v = a$ with probability $0 \leq \sigma_r \leq 1$ after receiving signal $s = r$.

Fixing the strategy of all members other than i we denote i's expected payoff from using strategy σ by:

$$\begin{aligned} E[U(\sigma, o, \omega)|s] &= \sigma_s \sum_{o \in \{a,r\}} \sum_{\omega \in \{A,R\}} \Pr[o, \omega|v = a, s] U(v = a, o, \omega) \\ &+ (1 - \sigma_s) \sum_{o \in \{a,r\}} \sum_{\omega \in \{A,R\}} \Pr[o, \omega|v = r, s] U(v = r, o, \omega), \end{aligned}$$

Throughout the analysis we rely on the concept of symmetric Bayesian Nash equilibrium.[6]

[5] In our use of the expressions "type I" and "type II" we follow what seems to be an emerging convention in the interdisciplinary literature that studies the FDA, rather than the standard definition used in hypothesis testing.

[6] Focusing on strategies that depend only on the players' types (in this case their signals) is a standard

3 Analysis

Fixing disesteem payoffs $\mathbf{k} = (k_1, k_2)$ and decision rule q, we denote by G^n the game with n players. The following expression represents an agent's relative expected payoff from voting to accept when all other agents play strategy σ:

$$\begin{aligned} & E[U(v=a,o,\omega)|s] - E[U(v=r,o,\omega)|s] \\ = & \sum_{o\in\{a,r\}} \sum_{\omega\in\{A,R\}} \Pr[o,\omega|v=a,s]U(a,o,\omega) \\ - & \sum_{o\in\{a,r\}} \sum_{\omega\in\{A,R\}} \Pr[o,\omega|v=r,s]U(r,o,\omega) \end{aligned}$$

An agent's vote only affects the committee outcome in the event that they are pivotal. Denoting this event by "Piv," the relative utility of voting a simplifies to:

$$\begin{aligned} & w\Pr[\text{Piv},\omega=A|s] - c\Pr[\text{Piv},\omega=R|s] \\ & - k_1\Pr[o=a,\omega=R|s] + k_2\Pr[o=a,\omega=A|s] \end{aligned} \tag{1}$$

Expression 1 illustrates the difference between the standard model of information aggregation and the model with disesteem payoffs. In the standard model, agents' votes are only payoff relevant when they are pivotal, represented by the first two terms of expression 1. In the model with disesteem payoffs, however, agents must also consider the probability of making personal type I and type II errors, represented by the second two terms of expression 1.

We first specify the general existence of a "babbling equilibrium," in which all agents vote to reject. Such equilibria generally exist in models of common-values voting (see Austen-Smith and Banks (1996)), since no single agent can affect the committee outcome if all other agents vote for the same option. Note, however, that a babbling equilibrium in which all agents vote to accept does not generally exist in our model, since agents are exposed to disesteem payoffs if the committee votes to accept.

Proposition 1 (Babbling Equilibrium)
For any q and $\mathbf{k}$, there exists N such that for $n > N$, $\sigma = (0, 0)$ is an equilibrium in game G^n.[7]

practice in Bayesian games. The justification in our case is that our players are ex-ante identical in every aspect but their labels and it is unappealing to let behavioral differences depend purely on payoff irrelevant characteristics (such as the labels). Restricting attention to symmetric strategies is also common in the voting literature when voting is simultaneous; see for example Palfrey and Rosenthal (1985) and Feddersen and Pesendorfer (1997).

[7] A sufficient condition is that $\lfloor Nq \rfloor \geq 1$ so that no agent can alter the commitee's decision by unilaterally changing his vote.

For the remainder of the analysis we restrict our attention to the existence and characterization of non-babbling equilibria.[8]

Next, we characterize large committee outcomes for $\mathbf{k} = (0, 0)$.

Proposition **2 (No disesteem payoffs)**
When $\mathbf{k} = (0, 0)$ *the decision of the committee converges almost surely to the state of the art for all* $q \in (0, 1)$ *as* n *approaches infinity.*

Proposition 2 states the analogous result to Feddersen and Pessendorfer's (1998) Proposition 3 for the SoA model (proved in the Supplementary Appendix as Corollary 2): in the absence of disesteem payoffs, regardless of q, decisions by large committees almost surely converge to the state of the art.

We now characterize the limit non-babbling equilibrium in the case in which committee members receive idiosyncratic payoffs for personal errors of type I and type II ($k_1 > 0$, $k_2 > 0$). The limit behavior of the committee as $n \to \infty$ can be generically characterized using the ratio of k_2/k_1.[9] That is, the absolute magnitude of the disesteem payoffs are irrelevant: no matter how small, the limit results of the model are determined by the relative magnitude of k_1 and k_2. Also, as shown by the following proposition, perfect information aggregation can only occur when $k_2/k_1 = e_2/(1 - e_2)$; for any value of k_2/k_1 other than this point, the committee decision, in the limit, is always biased towards either rejecting or accepting.

For notational ease, we let $\pi_1^n(\sigma) = \Pr[o = a|\tau = a]$ and $\pi_2^n(\sigma) = \Pr[o = a|\tau = r]$ – we explicitly denote the dependence of the probabilities on σ and n as we frequently need to appeal to them in our arguments. The following expressions are helpful in characterizing the limit equilibria:

$$\lambda = \frac{(1 - e_1)\Pr[\tau = r|s = r] + e_2 \Pr[\tau = a|s = r]}{(1 - e_2)\Pr[\tau = a|s = r] + e_1 \Pr[\tau = r|s = r]}$$

$$\overline{\pi}_2 = \left(\frac{k_2(1 - e_2) - k_1 e_2}{k_1(1 - e_1) - k_2 e_1} \right) \frac{p(\tau = a|s = r)}{p(\tau = r|s = r)}$$

In what follows, for each n, we denote an equilibrium strategy of G^n as σ^n, and for any sequence $\{\sigma^n\}$ we denote the limit of $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ as $n \to \infty$, when it exist, as (π_1, π_2).

[8]That is, statements such as "all equilibria of G^n...", should be read as "all equilibria of G^n, other than the babbling equilibrium,...".

[9]Besides the case in which $k_2/k_1 = e_2/(1 - e_2)$ where the asymptotic committee's behavior does depend on the actual magnitudes of k_1 and k_2, not just on the value of the ratio.

Proposition **3**

Assume $q \in (\varepsilon, 1-\varepsilon)$.[10] *For any sequence of non-babbling equilibria, the corresponding sequence $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ converges to:*

$$(\pi_1, \pi_2) = \begin{cases} (0,0) & \text{if } k_2/k_1 < e_2/(1-e_2) & \text{(i)} \\ (1, \overline{\pi}_2) & \text{if } k_2/k_1 \in (e_2/(1-e_2), \lambda] & \text{(ii)} \\ (1,1) & \text{if } k_2/k_1 > \lambda & \text{(iii)} \end{cases}$$

Moreover, there always exists a sequence of equilibria such that the corresponding sequence $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ converges to the specified values.[11]

Proposition 3 is the main result of our analysis and establishes the uniqueness, and for (ii) and (iii) the existence, of the non-babbling limit equilibrium (the formal proof of the proposition is given in the appendix). First, (i) demonstrates that the committee will always vote to reject the innovation as long as k_1 is large relative to k_2. An intuitive direct argument for this result stems from examining the right-hand side (RHS) and left-hand-side (LHS) of the following rearrangement of expert i's willingness to vote to accept the innovation upon receiving signal s, when all other members play according to σ,

$$\begin{aligned} &w\Pr[\text{Piv}, \omega = A|s] - c\Pr[\text{Piv}, \omega = R|s] + k_2\Pr[o = a, \omega = A|s] \\ &\geq k_1\Pr[o = a, \omega = R|s] \end{aligned} \tag{1”}$$

First, under any q-rule, the LHS converges to $k_2\Pr[o = a, \omega = A|s]$ as n approaches infinity, since the probability of influencing the committee decision approaches zero. Due to the state of the art layer, the RHS, while decreasing under some $\{\sigma_a, \sigma_r\}$, is always strictly bounded away from zero. Thus, for k_2/k_1 sufficiently small and n sufficiently large, experts will want to vote to reject the innovation independent of their signal.

Again, it is only the relative size of the disesteem payoffs that matters: it may be the case that rejecting all innovations with arbitrarily high probability (as the size of the committee increases) is the only equilibrium even for very small values of k_1. To be more precise, how small k_2/k_1 needs to be depends on the accuracy of the state of the art conditional on recommending approval; given both k_1 and k_2 strictly positive, the range over which rejecting the innovation is the unique limit outcome increases as the state of the art becomes less accurate.

[10] The results for $q \notin (\varepsilon, 1-\varepsilon)$ are largely analogous and are detailed in the supplementary appendix.

[11] The interval in (ii) is nonempty as long as at least one of e_1 and e_2 is strictly smaller than $1/2$, which by assumption is always the case throughout. Note that $k_2/k_1 = e_2/(1-e_2)$ is not explicitly covered by this proposition; however, this case turns out to be formally analogous to the game addressed in appendix 5.1 which implies that for certain parameter ranges, there exists an equilibrium sequence of $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ that converges to $(1, 0)$ (see Lemma 6 in section 5.1).

Similarly, (iii) demonstrates that if k_1 is small enough relative to k_2, a large committee essentially always accepts the innovation (outside of the babbling equilibrium). The intuition for this result is quite similar to (i): for k_2 large relative to k_1, as long as there is a positive probability of the committee passing the proposal, agents strictly prefer to vote for a to avoid k_2.

Case (ii) is less straightforward than the other two. For values of $k_2/k_1 > e_2/(1-e_2)$, conditional on the committee decision matching the state of the art, voting to reject is not a best response for an agent with a signal of r. Therefore, perfect information aggregation cannot be supported as an equilibrium in the limit; instead, agents with a signal of r mix between voting to accept and reject at a ratio such that $\pi_1^n(\sigma^n)$ equals 1 and $\pi_2^n(\sigma^n)$ is greater than zero, and their exposure to k_1 and k_2 is precisely balanced. The value of $\pi_2^n(\sigma^n)$ that makes agents with a signal of r indifferent between voting to reject and accept, $\bar{\pi}_2$, increases continuously from 0 to 1 as k_2/k_1 increases from $e_2/(1-e_2)$ to λ.

Note that there is no region of k_2/k_1 where $\pi_1 \in (0,1)$ and $\pi_2 = 0$. In fact, Proposition 3 implies a discontinuity between cases (i) and (ii) in the limit value of $\pi_1^n(\sigma^n)$, as $\pi_1 = 0$ for all $k_2/k_1 < e_2/(1-e_2)$ and $\pi_1 = 1$ for all $k_2/k_1 > e_2/(1-e_2)$. To see why there is no range of k_2/k_1 with $\pi_1 \in (0,1)$ and $\pi_2 = 0$, note that this would require agents with $s = a$ to mix between voting to reject and voting to accept. However, when $\pi_2 = 0$, an agent with $s = a$ can only be indifferent between $v = a$ and $v = r$ at the point $(k_2/k_1) = e_2/(1-e_2)$. That is, regardless of the value of π_1, when $k_2/k_1 < e_2/(1-e_2)$, it is strictly optimal to vote to reject for all agents, and when $k_2/k_1 > e_2/(1-e_2)$, it is strictly optimal for all agents with $s = a$ to vote to accept. This rules out an equilibrium of the form $\pi_1 \in (0,1)$ and $\pi_2 = 0$ (for $(k_2/k_1) \neq e_2/(1-e_2)$).

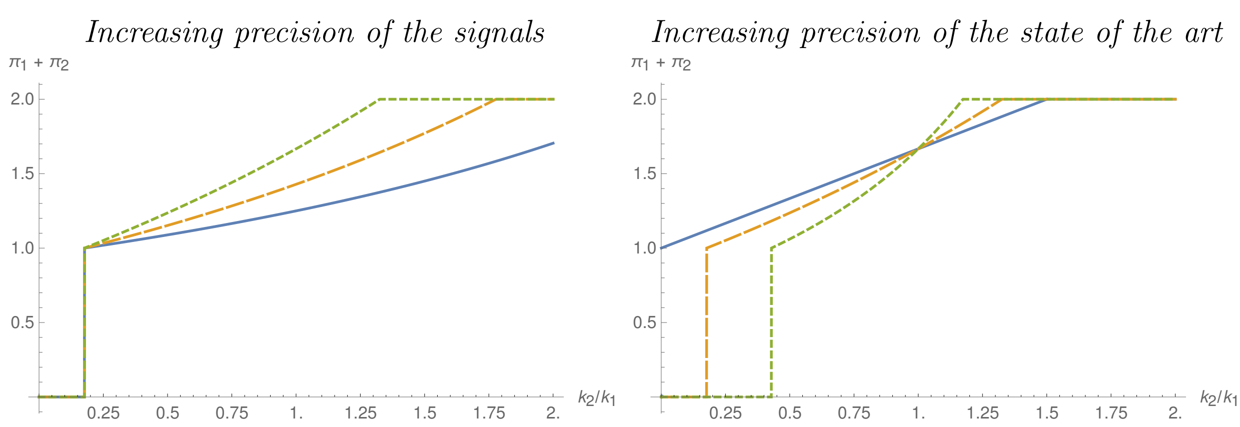

Figure 1: The left-hand graph shows $\pi_1 + \pi_2$ as a function of k_2/k_1 ($\pi_1 + \pi_2 = 1$ implies perfect information aggregation) for $e_2 = 0.15$ and $\epsilon = 0.4$ (small dashing), $\epsilon = 0.3$ (large dashing), and $\epsilon = 0.2$ (solid line). The right-hand graph shows $\pi_1 + \pi_2$ for $\epsilon = 0.4$ and $e_2 = 0.3$ (small dashing), $e_2 = 0.15$ (large dashing) and $e_2 = 0$ (solid line).

Next, Figure 1 provides some examples to illustrate the difference between changing the precision of the signal and changing the precision of the state of the art. The graphs show $\pi_1 + \pi_2$ as a function of k_2/k_1, where $\pi_1 + \pi_2 < 1$ corresponds to over-rejection, $\pi_1 + \pi_2 = 1$ corresponds to perfect information aggregation, and $\pi_1 + \pi_2 > 1$ corresponds to over-acceptance. The left-hand graph illustrates the effect of increasing ϵ while keeping e_2 constant.[12] Past the discontinuity, increasing ϵ decreases the probability that the committee incorrectly approves the innovation. The right-hand graph illustrates the effect of decreasing e_2 while keeping ϵ constant. Decreasing e_2 has a non-monotonic affect on the probability that the committee incorrectly approves the innovation – decreasing the probability of a type II error for small values of k_2/k_1, but increasing it for larger values.

Comparing the SoW and SoA models

Note that Proposition 3 implies a very significant difference between the SoW and the SoA models, which we emphasize in the following Corollary:

COROLLARY **1**
Under the SoW model, if disesteem payoffs are strictly positive, large committees always accept the innovation with too high a probability; i.e. $\pi_1 = 1, \pi_2 > 0$ *for any* $k_1, k_2 > 0$.

Under the SoA model, depending on the ratio of disesteem payoffs, large committees may accept innovations too often, or reject them too often; i.e. $\pi_1 = 1, \pi_2 > 0$ *for* k_2/k_1 *sufficiently large and* $\pi_1 = 0, \pi_2 = 0$ *for* k_2/k_1 *sufficiently small.*

Corollary 1, which follows as a direct application of Proposition 3 and is illustrated in Figure 2, shows that for all $k_1, k_2 > 0$, the SoW model predicts that a large committee of experts will always act rashly, accepting the innovation with too high a probability. It is only under the SoA model, for low k_2/k_1, that idiosyncratic payoffs lead to the over-caution of large committees of experts.

The difference between the two models is particularly stark when $k_1 > 0, k_2 = 0$.[13] In this case, under the SoW model, information aggregation can be sustained with disesteem payoffs since both the probability of being pivotal and the probability that the committee wrongly accepts the innovation approach zero for some σ. However, under the SoA model the probability that the committee wrongly accepts the innovation is bounded away from zero whenever $\Pr[o = A]$ is bounded away from zero. Therefore, the mechanism that sustains

[12] These examples, as well as the examples in Figure 2, set $e_1 = e_2$ and a 50-50 prior.

[13] Note that Proposition 3 does not cover the cases $k_1 > 0, k_2 = 0$, and $k_1 = 0, k_2 > 0$. For the SoA model (e_1, $e_2 > 0$), the proof of Proposition 3 extends to these corner cases. Under to SoW model, however, the case of $k_1 > 0, k_2 = 0$ requires a different approach, which we cover in the Appendix (see Proposition 4 in section 5.1).

information aggregation in the SoW model is absent in the SoA model. This difference holds in the limit as the SoA model approaches the SoW model (e_1 and e_2 approach zero), exposing a discontinuity in the standard model, where only a marginal deviation away from the SoW assumption changes equilibrium behavior from perfectly informative to babbling.

This discontinuity does not, however, exist when $k_1 = 0, k_2 > 0$, where both the SoA and SoW models predict that the committee will always vote to accept. This difference is due to an important, but subtle difference between disesteem payoffs for type I and type II errors. In particular, the possibility of facing k_1 depends on the event that the committee incorrectly accepts the innovation. In the SoW model, this event occurs with a vanishingly small probability as n grows, given that the votes by others are sufficiently informative.

In contrast, the possibility of facing k_2 depends on the event that the committee correctly accepts the innovation. In this case, assuming that the votes by others are sufficiently informative, this event happens with high probability in both the SoA and SoW models. Therefore, the incentives associated with k_2 matter for all n under both the SoA and SoW models. This implies that information aggregation will not be an equilibrium for $k_1 = 0, k_2 > 0$ for large n in either model, since $\Pr[o = a, \omega = A|s]$ remains strictly positive, which gives all agents a strict incentive to vote to accept.

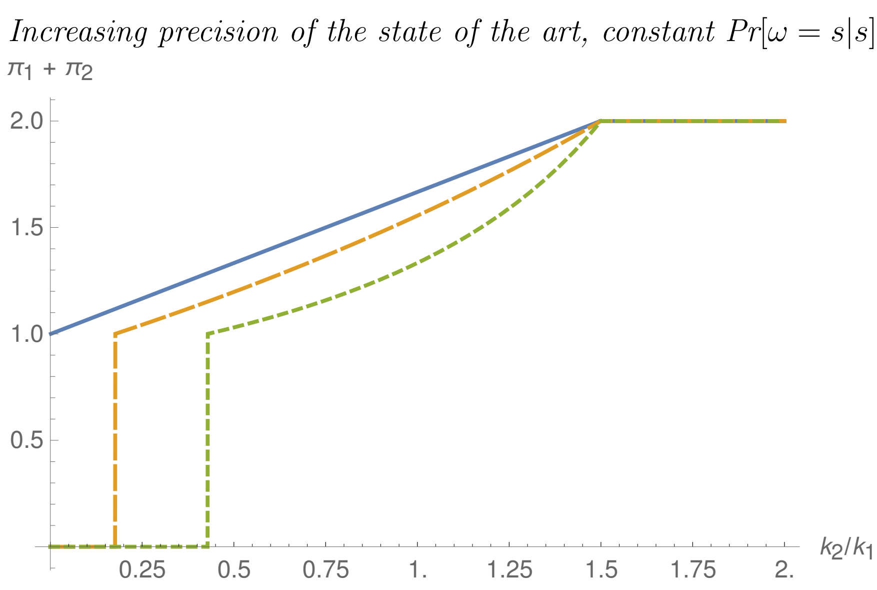

Figure 2: This graph shows $\pi_1 + \pi_2$ as a function of k_2/k_1 for $e_2 = 0$ (solid line), $e_2 = 0.15$ (large dashing), and $e_2 = 0.3$ (small dashing), adjusting ϵ so that $\Pr[\omega = s|s]$ stays constant at 0.6.

Lastly, Figure 2 gives a set of examples that illustrate the general difference between the SoW and SoA models. The graph shows $\pi_1 + \pi_2$ as a function of k_2/k_1 for $(\epsilon, e_2) = (0.6, 0), (0.63, 0.15), (0.75, 0.3)$. To provide a fair comparison of the SoW model ($e_2 = 0$)

and the SoA model ($e_2 > 0$), the values of ϵ are chosen such that $\Pr[\omega = s|s]$ is stays constant at 0.6. Figure 2 illustrates the stark discontinuity between the SoW and SoA models that occurs at $k_2/k_1 = 0$, with the SoW model predicting perfect information aggregation and the SoA model predicting that the committee will always reject. Moreover, relative to the SoW model, the SoA model predicts a lower probability of acceptance, with the difference between the two models increasing as more uncertainty is shifted from the signal to the state of the art.

4 Conclusion

In this paper, we detail the effect of disesteem payoffs on information aggregation in committees. We show that under the "state of the art" model of expertise, disesteem payoffs for personal type I errors can lead large committees to be over-cautious and reject new innovations as individual committee members seek to save face and avoid being blamed for a bad decision, while disesteem payoffs for personal type II errors can lead to large committees acting rashly as all committee members seek credit for approving good innovations.

Our paper also shows that the predictions of models of information aggregation can be sensitive to the standard assumption that experts' signals are independently distributed conditional on the state of the world: the standard model predicts that disesteem payoffs only cause rashness and never over-caution. The distinction between the two models is empirically relevant, since it is unlikely that the decision that aggregates all current knowledge perfectly identifies the true state of the world; that is, due to imperfect evidence, even the "best" decision might be wrong ex post. Additionally, the state of the art model in this paper implies a particular correlation structure between experts' signals, and the general implications of such correlation warrant further study.

Lastly, our paper shows that idiosyncratic payoffs can affect information aggregation even when they reinforce common payoffs. Specifically, idiosyncratic payoffs can distort decisions when they introduce asymmetry in payoffs. This asymmetry need not be large; we show here that even a marginal deviation from common payoffs can distort outcomes in large committees. Asymmetry can occur either due to informational asymmetry, e.g. when information regarding the adequacy of a drug is only revealed when the drug is passed, or if the saliency of individual votes vary with the committee outcome. One particularly relevant environment is a political setting, where idiosyncratic payoffs can be interpreted as changes in reelection probabilities due to the personal voting record of the politician. Voting records of politicians are heavily scrutinized in US legislatures, and the saliency of a particular representative's vote might condition on the legislative outcome. Therefore, an interesting area for future study is the effect of idiosyncratic payoffs on information aggregation in legislatures.

5 Appendix: Proofs

Proof of Proposition 1: Suppose every agent uses the strategy $\sigma = (0, 0)$. Choose any N such that $\lfloor Nq \rfloor \geq 1$. By this, none of the agents can unilaterally change the committee outcome and since the innovation is always rejected the disesteem payoffs are never realized. Hence, the agents are indifferent and we have an equilibrium. ▮*(Proposition 1)*

Proof of Proposition 2: Please see the supplementary appendix as it is the same argument as the one used in the analogous result to Feddersen and Pessendorfer's (1998) Proposition 3.

Proof of Proposition 3:
The proof is divided into three parts. The first part of the proof establishes some useful preliminary lemmas. We begin with the statement, in Lemma 1, of a standard result on the uniform convergence of points of mass in binomial distributions which underlies our main arguments. Lemma 2 shows that for large enough n the function describing agent i's willingness to reject the innovation conditional on his signal, in game G^n, can be approximated by a function that does not include any terms involving the probability of the event that i is pivotal. Importantly, this approximation has a lower bound that is independent of the strategies of the other agents. Lemma 3 and Lemma 4 show that for large enough n, equilibria of the games G^n must have a very specific shape. These restrictions make it considerably easier to characterize equilibria.

In the second part of the proof we establish the existence of sequences of equilibria with conditional acceptance probabilities that converge to the values specified in the statement of the proposition. The method involves relying on the much simpler approximate expressions for the willingness to reject provided by Lemma 2. The main argument in this section is embodied in Lemma 5. Finally, note that $k_2/k_1 = e_2/(1 - e_2)$ is not explicitly covered by proposition 3 since a different method of analysis is required for this case. However, we do address this case in Lemma 6 in section 5.1, and show that for certain parameter ranges there exists an equilibrium sequence that converges to perfect information aggregation.

The third part of the proof shows that the sequences constructed in the first part are essentially unique in the sense that any sequence of non-babbling equilibria must yield the same limiting conditional acceptance probabilities.

Part 1 (Preliminary Lemmas)

As seen throughout the paper, the rate at which the probability of being pivotal converges to 0 as the size of committees grows plays a crucial role in the analysis behind many of our results. From the perspective of any agent, the probability of the event that he is pivotal corresponds to there being a precise number of other agents, $\lfloor nq \rfloor$, (conditional on the state of the art) voting to accept. From his perspective, conditional on the state of the art, this number is binomially distributed, and the success probability governing such a distribution is given by the probability that a randomly chosen agent in the committee votes to accept. Since this is an equilibrium object the following Lemma, which shows that for any given q, the probability of being pivotal converges uniformly to 0 as $n \to \infty$ (uniformly in the probability of voting to accept) is particularly useful. This is a standard property of sequences of binomial pmfs, but since it plays such an important role in our analysis we state it below.

LEMMA **1 (Convergence of binomial points of mass)**
The set $\{\binom{n-1}{\lfloor nq \rfloor} p^{\lfloor nq \rfloor}(1-p)^{n-1-\lfloor nq \rfloor} : 0 \leq p \leq 1\}$ *is bounded above by a function* $f(n)$ *such that* $\lim_{n\to\infty} f(n) \to 0$.

Proof of Lemma 1:
It follows by applying Stirling's formula to establish an upper bound for the set $\{\binom{n-1}{\lfloor nq \rfloor} p^{\lfloor nq \rfloor}(1-p)^{n-1-\lfloor nq \rfloor} : 0 < p < 1\}$ and showing that this upper bound converges to 0. ▮*(Lemma 1)*

The proof of our main result, Proposition 3, relies on some properties of the willingness to vote to reject as $n \to \infty$ that we establish in the following lemmas. For the following, this function of $\pi = (\pi_1, \pi_2)$ will be useful:

$$\begin{aligned}\mathbf{R}_s(\pi) = &\pi_1\big(k_1\text{Pr}[\omega = R|\tau = a] - k_2\text{Pr}[\omega = A|\tau = a]\big)\text{Pr}[\tau = a|s] + \\ &\pi_2\big(k_1\text{Pr}[\omega = R|\tau = r] - k_2\text{Pr}[\omega = A|\tau = r]\big)\text{Pr}[\tau = r|s],\end{aligned}$$

since, as the first Lemma in the series shows, the willingness to vote to reject when observing signal s gets arbitrarily close to $\mathbf{R}_s(\pi)$ for large n.

LEMMA **2**
For any $h > 0$, *there exists* N *such that for all* $n > N$, *the willingness to vote to reject* $R_s(n, \sigma)$ *is within* h *of* $\mathbf{R}_s(\pi_1^n(\sigma), \pi_2^n(\sigma))$, *where* $\pi_1^n(\sigma) = Pr[o = A|\tau = a]$ *and* $\pi_2^n(\sigma) = Pr[o = A|\tau = r]$.[14]

Proof: For n finite, the willingness to vote to reject can be rewritten as follows:

$$\begin{aligned}R_s(n, \sigma) = &- w\text{Pr}[\text{Piv}, \omega = A|s] + C\text{Pr}[\text{Piv}, \omega = R|s] + k_1\text{Pr}[\text{Piv}, \omega = R|s] \\ &+ k_1\text{Pr}[o = a, \neg\text{Piv}, \omega = R|s] - k_2\text{Pr}[o = a, \neg\text{Piv}, \omega = A|s],\end{aligned}$$

where $\neg$Piv indicates the event that i is not pivotal. This in turn can be written as:

$$F_s'(\text{Piv}, \sigma, n) + k_1\text{Pr}[o = a, \neg\text{Piv}, \omega = R|s] - k_2\text{Pr}[o = a, \neg\text{Piv}, \omega = A|s],$$

[14] For notational ease, when there is no risk of confusion we do not explicitly denote the dependence of the probability terms on n.

where $F'_s(\text{Piv}, \sigma, n)$ gathers the terms that involve the event that i is pivotal.

With some additional algebra, the above expression can be shown to equal:

$$F'_s(\text{Piv}, \sigma, n)+[k_1\Pr[\omega = R|\tau = a] - k_2\Pr[\omega = A|\tau = a]]\Pr[\tau = a|s]\Pr[o = a, \neg\text{Piv}|\tau = a]+ [k_1\Pr[\omega = R|\tau = r] - k_2\Pr[\omega = A|\tau = r]]\Pr[\tau = r|s]\Pr[o = a, \neg\text{Piv}|\tau = r].$$

Since $\Pr[o = a, \neg\text{Piv}|\tau = a] = \pi_1^n(\sigma) - \Pr[o = a, \text{Piv}|\tau = a]$ and $\Pr[o = a, \neg\text{Piv}|\tau = r] = \pi_2^n(\sigma) - \Pr[o = a, \text{Piv}|\tau = r]$, we can substitute these terms into the above equation and rearrange to get:

$$R_s(n, \sigma) = \pi_1^n(\sigma)[k_1e_2-k_2(1-e_2)]\Pr[\tau = a|s]+\pi_2^n(\sigma)[k_1(1-e_1)-k_2e_1]\Pr[\tau = r|s]+F_s(\text{Piv}, \sigma, n),$$

where $F_s(\text{Piv}, \sigma, n)$ captures all payoffs associated with the pivotal event. Note that since the probability of being pivotal approaches zero uniformly with respect to σ (see Lemma 1), there exists a function $m_s(n)$ converging to 0 as $n \to \infty$ such that for all σ we have $|F_s(\text{Piv}, \sigma, n)| < m_s(n)$. Let $m(n) = max\{m_a(n), m_r(n)\}$ and consider N such that $m(n) < h$ for all $n > N$. Then we have that for all $n > N$:

$$|R_s(n, \sigma) - \mathbf{R}_s(\pi_1^n(\sigma), \pi_2^n(\sigma))| < h$$

∎*(Lemma 2)*

Next, we prove that equilibrium strategies can only take a certain form. This fact greatly simplifies the task of characterizing the limit behavior of the committee:

Lemma **3**
If $k_2/k_1 > e_2/(1 - e_2)$, $k_2/k_1 \neq (1 - e_1)/e_1$, then for any $\delta > 0$ there exists N such that for all $n > N$, if there is an equilibrium of G^n such that either $\pi_1^n(\sigma^n) > \delta$ or $\pi_2^n(\sigma^n) > \delta$, then $\sigma_a^n = 1$ and $\sigma_r^n \leq 1$ or $\sigma_a^n \geq 0$ and $\sigma_r^n = 0$.

Proof of Lemma 3:
First, consider the case of $k_2/k_1 \in (e_2/(1 - e_2), (1 - e_1)/e_1)$. Take $h > 0$ small enough such that the following inequality holds:

$$\begin{aligned}&\mathbf{R}_r(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n)) - \mathbf{R}_a(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\\ &=\pi_1^n(\sigma^n)\big[\big(k_1e_2 - k_2(1 - e_2)\big)(\Pr[\tau = a|s = r] - \Pr[\tau = a|s = a])\big]\\ &\quad + \pi_2^n(\sigma^n)\big[\big(k_1(1 - e_1) - k_2e_1\big)(\Pr[\tau = r|s = r] - \Pr[\tau = r|s = a])\big] > 2h,\end{aligned}$$

where h is well-defined since both terms within brackets are strictly positive for $k_2/k_1 \in (e_2/(1 - e_2), (1 - e_1)/e_1)$, and either $\pi_1^n(\sigma^n)$ or $\pi_2^n(\sigma^n)$ is strictly greater than δ.

Take N large enough such that Lemma 2 holds for h, then we have that for $n > N$:

$$\mathbf{R}_r(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n)) - \mathbf{R}_a(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n)) > 2h \Rightarrow R_r(n, \sigma) - R_a(n, \sigma) > 0$$

So any equilibrium of G^n such that either $\pi_1^n(\sigma^n) > \delta$ or $\pi_2^n(\sigma^n) > \delta$ must be of the form $\sigma_a^n = 1$ and $\sigma_r^n \leq 1$ or $\sigma_a^n \geq 0$ and $\sigma_r^n = 0$.

Next, we consider the case of $k_2/k_1 > (1-e_1)/e_1$. When $k_2/k_1 > (1-e_1)/e_1$, then $\mathbf{R}_s(\pi) < 0$ for all $\pi \neq (0,0)$. Therefore, by Lemma 2 for n large enough the only possible equilibria are $\sigma = (0,0), (1,1)$. ▮*(Lemma 3)*

Lemma 4 allows us to further narrow the candidate non-babbling equilibria in the case of $\varepsilon < q < 1 - \varepsilon$.

Lemma **4**
If $\varepsilon < q < 1-\varepsilon$ and if $k_2/k_1 > e_2/(1-e_2)$, $k_2/k_1 \neq (1-e_1)/e_1$, then for any $\delta > 0$, there exists N such that for all $n > N$, if there is an equilibrium of G^n such that either $\pi_1^n(\sigma^n) > \delta$ or $\pi_2^n(\sigma^n) > \delta$, then $\sigma_a^n = 1$ and $\sigma_r^n \leq 1$.

Proof of Lemma 4:
By contradiction, assume that for all N, there is $n > N$ such that there exists an equilibrium, $\hat{\sigma}^n$, such that $\hat{\sigma}_r^n = 0$ and $\hat{\sigma}_a^n \geq 0$, and $\pi_1^n(\hat{\sigma}^n) > \delta$ or $\pi_2^n(\hat{\sigma}^n) > \delta$. Consider N_1 large enough so that Lemma 2 holds for h and so that we can apply Lemma 3 for all $n > N_1$. Given $\sigma^n(r) = 0$ and $\sigma^n(a) \geq 0$, then there exists $N > N_1$ such that for all $n > N$, $\pi_2^n(\sigma^n) < \delta$ and $\pi_2^n(\sigma^n)\big(k_1(1-e_1) - k_2 e_1\big)\Pr[\tau = r|s = r] < -\delta\big(k_1 e_2 - k_2(1-e_2)\big)\Pr[\tau = a|s = r] - h$ for some small h. This implies:

$$\begin{aligned}\mathbf{R}_r(\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n)) + h = \pi_1^n(\hat{\sigma}^n)\big(k_1 e_2 - k_2(1-e_2)\big)\Pr[\tau = a|s = r] \\ + \pi_2^n(\hat{\sigma}^n)\big(k_1(1-e_1) - k_2 e_1\big)\Pr[\tau = r|s = r] + h < 0,\end{aligned}$$

since if $\pi_2^n(\hat{\sigma}^n) < \delta$ it must be the case that $\pi_1^n(\hat{\sigma}^n) > \delta$.

However, Lemma 2 implies

$$R_r(n, \hat{\sigma}) < \mathbf{R}_r(\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n)) + h < 0,$$

which contradicts $\hat{\sigma}_r^n = 0$. ▮*(Lemma 4)*

Part 2 (Existence)

We now prove the statement on existence of the proposition. That is, we prove that when k_2/k_1 lies in each of the sets specified below, then there exists a sequence of equilibria $\{\sigma^n\}$ such that the corresponding sequence of conditional acceptance probabilities $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ converges to the specified values:

(i)	$(0,0)$	if $k_2/k_1 < e_2/(1-e_2)$
(ii)	$(1,\bar{\pi}_2)$	if $k_2/k_1 \in (e_2/(1-e_2),\lambda]$
(iii)	$(1,1)$	if $k_2/k_1 > \lambda$

Note that (i) follows trivially from the existence of a babbling equilibrium at $\sigma_a = \sigma_b = 0$. Also, we consider the cases of $\frac{k_2}{k_1} = \frac{1-e_1}{e_1}$, $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$ at the end of the proof, since these cases require special treatment.

For $\frac{k_2}{k_1} > \frac{e_2}{1-e_2}$, $\frac{k_2}{k_1} \neq \frac{1-e_1}{e_1}$, Lemma 3 allows us to constrain the analysis to $\sigma_r = 0, 0 \leq \sigma_a \leq 1$, and $0 \leq \sigma_r \leq 1, \sigma_a = 1$, allowing us to define equilibria of the game using the same willingness to vote to reject function used for the main analysis:

$$R(n,z) = \begin{cases} R_a(n,(z,0)) & \text{if } z \leq 1 \\ R_r(n,(1,z-1)) & \text{if } z > 1 \end{cases}$$

Where $z = \sigma_a + \sigma_r$.[15] Also, since either $\pi_1^\infty(z) \in [0,1]$, $\pi_2^\infty(z) = 0$, or $\pi_1^\infty(z) = 1$, $\pi_2^\infty(z) \in [0,1]$, it will be useful to define $\mathbf{R}(\mathit{\Pi})$ as a function of $\mathit{\Pi} = \pi_1 + \pi_2$ over this subset of (π_1,π_2). Formally:

$$\mathbf{R}(\mathit{\Pi}) = \begin{cases} \mathbf{R}(\pi_1,0) & \text{if } \pi_2 = 0 \\ \mathbf{R}(1,\pi_2)) & \text{if } \pi_1 = 1 \end{cases}$$

From this expression we derive the bounds of λ and $e_2/(1-e_2)$, as well as $\bar{\pi}_2$. Specifically, λ is the value of k_2/k_1 that solves $\mathbf{R}(2) = 0$, implying that $\mathbf{R}(2) < 0$ for values of $k_2/k_1 > \lambda$. Similarly, $e_2/(1-e_2)$ solves $\mathbf{R}(0) = 0$, and $\bar{\pi}_2$ solves $\mathbf{R}(1+\bar{\pi}_2) = 0$ for $k_2/k_1 \in (e_2/(1-e_2),\lambda)$.

Moreover, this notation allows us to prove the following lemma:[16]

[15] What is crucial here, is that when characterizing equilibria in mixed strategies we only need to verify the indifference of the agent between accepting and rejecting, after observing the signal under which he mixes. Lemma 3 then implies the optimality of his behavior under the other signal when n is large enough.

[16] We are interested in the roots of $\mathbf{R}(\mathit{\Pi})$ to the extent that they correspond to limit points of acceptance probabilities in sequences of equilibria of the finite games. The fact that π that do not have this shape are not candidate limit point stems from the informativeness of the signals with respect to the state of the art ($\varepsilon < \frac{1}{2}$). This assumption along with the shape of the equilibria, imply that from the perspective of an observer the probability μ_a^n that a randomly chosen agent votes to accept conditional on $\tau = a$ is always strictly higher than conditional on $\tau = r$ (μ_r^n) -as long as $\sigma^n(a)$ and $\sigma^n(r)$ are not both 0 or both 1-. By the law of large numbers, $\Pr[o = A|t] \in (0,1)$ converges to an interior point of $[0,1]$ only if $\mu_t^n \to q$, but it can't be the case that both μ_r^n and μ_a^n are converging to an interior point of $[0,1]$.

Lemma 5

(i) Given (π_1^*, π_2^*) *such that* $\Pi^* \in (0,1)$ *and* $\mathbf{R}(\Pi^*) = 0$*, there exists a sequence of equilibria* $\{\sigma^n\}$ *such that as* $n \to \infty$ *the corresponding sequence of* $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ *converges to* (π_1^*, π_2^*) *if* $\mathbf{R}(\Pi^* - h) < 0$ *and* $\mathbf{R}(\Pi^* + h) > 0$ *for all* $h \in (0, \bar{h})$ *for some* $\bar{h} > 0$*.*

(ii) If $\mathbf{R}(2) \leq 0$*, there exists a sequence of equilibria* $\{\sigma^n\}$ *such that as* $n \to \infty$ *the corresponding sequence of* $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ *converges to* $\pi = (1,1)$*.*

Proof of Lemma 5: We use Lemma 2 and the following expression,

$$R_s(n,\sigma) = \pi_1^n(\sigma)[k_1 e_2 - k_2(1-e_2)]\Pr[\tau = a|s] + \pi_2^n(\sigma)[k_1(1-e_1) - k_2 e_1]\Pr[\tau = r|s] + F_s(\text{Piv}, \sigma, n),$$

to prove (i) by first proving that for n large, there exists an equilibrium within an h-neighborhood of (π_1^*, π_2^*) for each n, and then proving that a sequence of equilibria within the h-neighborhood of (π_1^*, π_2^*) converges to (π_1^*, π_2^*).

Take a sequence of $\{z_1^n\}$, such that $(\pi_1^n(z_1^n), \pi_2^n(z_1^n))$ converges to $(\Pi^* + h)$, and $\{z_2^n\}$, such that $(\pi_1^n(z_2^n), \pi_2^n(z_2^n))$ converges to $(\Pi^* - h)$. Such sequences exist since π_1, π_2 are continuous in z. Moreover, for all $n > N^+$, for some N^+, $R(n, z_1^n) > 0$ by Lemma 2. Similarly, for all $n > N^-$, for some N^-, $R(n, z_2^n) < 0$. This implies that for each value of n greater than N, where $N = \max\{N^+, N^-\}$, there exists an equilibrium value z^*, such that $R(n, z^*) = 0$, since $R(n, z_2^n) < 0 < R(n, z_1^n)$ and $R(n, z)$ is continuous in z. This proves the existence of a sequence of equilibria within a h-neighborhood of (π_1^*, π_2^*) for $n > N$.

Take a sequence of equilibria, z^{n*}, such that the corresponding sequence of $\{\pi_1^n(z^{n*}), \pi_2^n(z^{n*})\}$ is within a h-neighborhood of (π_1^*, π_2^*) for $n > N$, where h is small enough so that the neighborhood does not include $(0,0)$. By construction $R(n, z^{*,n}) = 0$ all along the sequence, and furthermore $F_s(\text{Piv}, \sigma, n) \to 0$. Therefore, $\pi_1^n(z^n)[k_2 e_2 - k_1(1-e_2)]\Pr[\tau = a|s] + \pi_1^n(z^n)[k_2 e_1 - k_1(1-e_1)]\Pr[\tau = r|s]$ must converge to 0. This can only happen if (1) $(\pi_1^n(z^{n*}), \pi_2^n(z^{n*}))$ are converging to (π_1^*, π_2^*); (2) $(\pi_1^n(z^{n*}), \pi_2^n(z^{n*}))$ are converging to $(0,0)$; or (3) $(\pi_1^n(z^{n*}), \pi_2^n(z^{n*}))$ is alternating between a neighborhood of (π_1^*, π_2^*) and a neighborhood of $(0,0)$. However, since $(\pi_1^n(z^{n*}), \pi_2^n(z^{n*}))$ is bounded from $(0,0)$ for $n > N$ by construction, (2) and (3) are impossible.

The proof of (ii) follows a similar logic. First, if $\mathbf{R}(2) = 0$, then a sequence of equilibria such that the corresponding $\{(\pi_1^n(z^{n*}), \pi_2^n(z^{n*}))\}$ converge to $(1,1)$ exists by the same argument as above since $\mathbf{R}(\Pi^* - h) < 0$ for all h.

For $\mathbf{R}(2) < 0$, take $h > 0$, but small enough that $\mathbf{R}(2) + h < 0$. Next, take N large enough such that $|F(\text{Piv}, z, n)| < h$ for all $n > N$. This implies that for $n > N$, $z = 2$ is an equilibrium since:

$$R_s(n,2) = \mathbf{R}_s(2) + F(\text{Piv}, z, n) < \mathbf{R}(2) + h < 0.$$

This in turn implies that a sequence of equilibria exist such that $(\pi_1^n(z^{n*}), \pi_2^n(z^{n*})) = (1, 1)$ is an equilibrium for all $n > N$. ▮*(Lemma 5)*

The proof of (ii) and (iii) now follow directly from Lemma 5: For (ii), $\mathbf{R}(1 + \bar{\pi}_2) = 0$ and $\mathbf{R}(1 + \bar{\pi}_2 - h) < 0 < \mathbf{R}(1 + \bar{\pi}_2 + h)$ for all $h \in (0, \bar{h})$ for $\bar{h}$ small. For (iii), $\mathbf{R}(2) \leq 0$. For $\frac{k_2}{k_1} = \frac{1-e_1}{e_1}$, Lemma 5 does not apply; however, since $\mathbf{R}_s(1, 1) < 0$ it follows that an equilibrium with $(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))$ close to $(1, 1)$ exists for n large.

Part 3 (Uniqueness)

We now prove the uniqueness of the limits of conditional probabilities associated to sequences of non-babbling equilibria, as discussed at the beginning of the statement of Proposition 3. That is, we prove that the sequence of conditional acceptance probabilities $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ associated to any given sequence $\{\sigma^n\}$ of non-babbling equilibria converges to:

(i) $(0, 0)$ if $k_2/k_1 < e_2/(1 - e_2)$
(ii) $(1, \bar{\pi}_2)$ if $k_2/k_1 \in (e_2/(1 - e_2), \lambda]$
(iii) $(1, 1)$ if $k_2/k_1 > \lambda$

We proceed by contradiction, assuming that there exists δ such that for all N, there is $n > N$ such that there exists an equilibrium $\hat{\sigma}^n$ such that either $\pi_1^n(\hat{\sigma}^n)$ or $\pi_2^n(\hat{\sigma}^n)$ is more than δ away from the points specified in the proposition. The details of the proof differ between the individual cases. However, for each case, we make use of the fact that by Lemma 2, given σ^n, the willingness to vote to reject, $R_s(n, \sigma^n)$, will be close to:

$$\mathbf{R}_s(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n)) = \pi_1^n(\sigma^n)\big(k_1e_2 - k_2(1 - e_2)\big)\Pr[\tau = a|s] + \pi_2^n(\sigma^n)\big(k_1(1 - e_1) - k_2e_1\big)\Pr[\tau = r|s],$$

and proceed by showing that $\hat{\sigma}^n$ cannot be an equilibrium for n large.

Case (i) $\left(\frac{k_2}{k_1} < \frac{e_2}{1-e_2}\right)$*:* Here we consider $\hat{\sigma}^n$ such that either $\pi_1^n(\hat{\sigma}^n)$ or $\pi_2^n(\hat{\sigma}^n)$ is more than δ away from 0. Take $h > 0$ such that $min\{\delta(k_1e_2 - k_2(1 - e_2))\Pr[\tau = a|s], \delta(k_1(1 - e_1) - k_2e_1)\Pr[\tau = a|s]\} > h$ for each signal $s = a$ and $s = r$ (such an h exists since both expressions within the brackets are strictly positive given $k_2/k_1 < e_2/(1 - e_2) < (1 - e_1)/e_1$).

By assumption, there exists $\hat{\sigma}^n$ such that either $\pi_1^n(\hat{\sigma}^n) > \delta$ or $\pi_2^n(\hat{\sigma}^n) > \delta$ and, by Lemma 2, such that $R_s(n, \hat{\sigma}^n)$ is within h of $\mathbf{R}_s(\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n))$. Therefore:

$$\begin{aligned} R_s(n, \hat{\sigma}^n) >& \pi_1^n(\hat{\sigma}^n)\big(k_1e_2 - k_2(1 - e_2)\big)\Pr[\tau = a|s = r] + \pi_2^n(\hat{\sigma}^n)\big(k_1(1 - e_1) - k_2e_1\big)\Pr[\tau = r|s] - h \\ >& \min\{\delta(k_1e_2 - k_2(1 - e_2))\Pr[\tau = a|s], \delta(k_1(1 - e_1) - k_2e_1)\Pr[\tau = a|s]\} - h > 0. \end{aligned}$$

So it follows that the unique best response of an agent is to vote to reject under both signals, and thereby $\hat{\sigma}^n$ with corresponding $\pi_1^n(\hat{\sigma}^n)$ cannot be an equilibrium of G^n.

Case (ii) $\frac{k_2}{k_1} \in \left(\frac{e_2}{1-e_2}, \lambda\right]$*:* Here we consider $\hat{\sigma}^n$ such that $(\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n))$ is more than δ away from $(0,0)$ or $(1, \bar{\pi}_2)$, in either dimension.

First notice that within the stated range of k_2/k_1, the unique interior crossing of $\mathbf{R}(\mathit{\Pi})$ is at $(1+\bar{\pi}_2)$, which implies that for some small enough h, there exists $x > 0$ such that if $\pi_1^n(\sigma^n) > 1-x$ and $|\pi_2^n(\sigma^n) - \bar{\pi}_2| > \delta$ then $|\mathbf{R}_r(\infty, (\pi_1^n(\sigma^n), \pi_2^n(\sigma^n)))| > h$. Take h small enough so that this relationship holds, and so that $\mathbf{R}_r(\infty,(1,1)) > h$ and $\mathbf{R}_r(\infty,(1,0)) < -h$ (these two expressions are strictly positive and negative (respectively) for the range of k_2/k_1 considered).

Next, pick N large enough so that (1) Lemma 4 holds for δ, (2) Lemma 2 holds for h, (3) $\pi_1^n(\hat{\sigma}^n) > \max\{1-\delta, 1-x\}$, and (4) $R_r(n,(1,0)))$ is within h of $\mathbf{R}_r(\infty,(1,0))$ (which follows from the fact that $q \in (\varepsilon, 1-\varepsilon)$ implies $\lim_{n\to\infty}(\pi_1^n(1,0), \pi_2^n(1,0)) = (1,0)$). It follows that for the conjectured equilibrium $\hat{\sigma}^n$, $R_a(n,\hat{\sigma}^n) \neq 0$ and $R_r(n,\hat{\sigma}^n) \neq 0$ since $\mathbf{R}_s((\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n)))$ are at least h away from 0. So both $\hat{\sigma}_a^n$ and $\hat{\sigma}_r^n$ must be extreme points of $[0,1]$, which implies that the only candidates for $\hat{\sigma}^n$ are $(0,0)$, $(1,0)$ and $(1,1)$ (pure strategies).

However, $\hat{\sigma}^n$ cannot equal $(0,0)$, as this implies $(\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n)) = (0,0)$. Nor can $\hat{\sigma}^n$ equal $(1,1)$, as this requires $R_a(n,\sigma^n) < 0$ and $R_r(n,\sigma^n) < 0$ and implies $(\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n)) = (1,1)$, which cannot be true since $\mathbf{R}_r(\infty,(1,1)) > h$ implies $R_r(n,(1,1)) > 0$. Finally, $\hat{\sigma}^n$ cannot equal $(1,0)$, since this requires $R_a(n,\sigma^n) < 0$ and $R_r(n,\sigma^n) > 0$, which cannot be true since $\mathbf{R}_r(\infty,(1,0)) < -h$ and $R_r(n,(1,0)))$ is within h of $\mathbf{R}_r(\infty,(1,0))$, implying $R_r(n,(1,0))) < 0$. This chain of arguments contradicts that σ^n is an equilibrium of G^n.

Case (iii) $\frac{k_2}{k_1} > \lambda$*:* Here we consider $\hat{\sigma}^n$ such that either $\pi_1^n(\hat{\sigma}^n)$ or $\pi_2^n(\hat{\sigma}^n)$ is more than δ away from 1 and 0.

Note that since $\mathbf{R}(\mathit{\Pi}) < 0$ for all $\mathit{\Pi} \in (0,1]$, $R_a(n,\hat{\sigma}^n) \neq 0$ and $R_r(n,\hat{\sigma}^n) \neq 0$ for large enough n by the same argument as in Case (ii). Therefore, $\hat{\sigma}^n$ must be a pure strategy equilibrium. However, as above, $\hat{\sigma}^n$ cannot equal $(0,0)$ or $(1,1)$ since this implies $(\pi_1^n(\hat{\sigma}^n), \pi_2^n(\hat{\sigma}^n))$ equals $(0,0)$ or $(1,1)$ (respectively). Also, $\hat{\sigma}^n$ cannot equal $(1,0)$ for n large enough since, as above, $\mathbf{R}_r(\infty,(1,0)) < -h$.

Finally, note that when $k_2/k_1 = (1-e_1)/e_1$, the same argument as above demonstrates that all equilibria with $\pi_1^n(\sigma_n) >> 0$ must be within δ of $(1,1)$ for n large. However, since $\mathbf{R}(0,\pi_2) = 0$ at this point, Lemma's 3 and 4 do not apply, and we cannot exclude the existence of equilibria close to $(0,\pi_2)$. ▮*(Proposition 3)*

5.1 Conditions for information aggregation with disesteem payoffs

Here we consider two related cases where perfect information aggregation can be supported with disesteem payoffs.

SoW model, $k_1 > 0$, $k2 = 0$:
The proof of Proposition 3 does not cover the case of $k_1 > 0$, $k_2 = 0$ under the SoW model ($e_1 = e_2 = 0$). Therefore, we present the following proposition, which provides a sufficient condition for perfect information aggregation under the SoW model:

PROPOSITION 4 **(Payoffs for Type I errors, SoW: Information aggregation)**
Assume that $e_1 = e_2 = 0$, $\frac{w}{c} \in \left(\frac{\varepsilon^2(1-p_A)}{(1-\varepsilon)^2 p_A}, \frac{(1-p_A)}{p_A}\right)$, *and*

$$k_1 < \frac{p_A w \left(\frac{(1-\varepsilon)^2}{\varepsilon}\right) - c\varepsilon(1-p_A)}{\varepsilon(1-p_A)\frac{1-\varepsilon}{1-2\varepsilon}} \tag{2}$$

If $q = \frac{1}{2}$, *then* $\sigma_a = 1$, $\sigma_r = 0$ *is an equilibrium for all sufficiently large* n. *In particular, this implies that the Condorcet Theorem holds for all* k_1 *satisfying the inequality above.*[17]

With $k_1 > 0$ and $k_2 = 0$, the relative payoff from voting to accept, expression 1, simplifies to:

$$w\Pr[\text{Piv}, \omega = A|s] - c\Pr[\text{Piv}, \omega = R|s] - k_1 \Pr[o = a, \omega = R|s]. \tag{1'}$$

The intuition for the proof of Proposition 4 (see the supplementary appendix for the formal proof) is that, given $\sigma_a = 1$, $\sigma_r = 0$, the probability that the committee makes the wrong decision converges to 0 at the same rate as the probability that a given agent is pivotal, and thus the ratio of the two probabilities, $\Pr[\text{Piv}|\omega = A]/\Pr[o = A|\omega = R]$, approaches a strictly positive constant. Therefore, when k_1 is sufficiently small relative to w, the relative benefit of voting to accept given a signal of accept outweighs the exposure to the disesteem payoff in large committees (expression 1' is positive in the limit for $s = a$, and negative for $s = r$), and truthful voting is supported in equilibrium.

SoA model, $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$:
Here we show that when $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$, information aggregation can be supported in the limit for certain parameter ranges. Notice that for this value of k_2/k_1, $\mathbf{R}_s(\pi_1, \pi_2)$ is equal to:

$$\mathbf{R}_s(\pi_1, \pi_2) = \pi_2\big(k_1(1-e_1) - k_2 e_1\big)\Pr[\tau = r|s],$$

[17]Note that this is a sufficient, but not necessary, condition. In particular, this proposition specifies conditions under which voting is truthful ($\sigma_a = 1$, $\sigma_r = 0$) given a majority rule and k_1 positive, a stronger condition than is needed for the Condorcet Theorem to hold. Also, if the disesteem payoff is "diluted" as n grows, then the condition on k_1 is not restrictive. We discuss this, and the robustness of the following SoA result to dilution, in a working version of the paper (see Midjord et al. (2014)).

which suggests that information may be aggregated in the limit, since information aggregation implies $\lim_{n\to\infty} \pi_2^n(\sigma^n) = 0$. In fact, we show below that the case of $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$ is analogous to the SoW model, which implies that information aggregation in reached in the limit for the comparable parameters detailed in Proposition 4.

Lemma **6**
When $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$, then there exist parameter values such that $\{(\pi_1^n(\sigma^n), \pi_2^n(\sigma^n))\}$ converges to $(1,0)$.

Proof of Lemma 6:
First, we show that the case of $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$ is analogous to the SoW model. Under parameters w, c, k_1, k_2, ε, p_A, e_1 and e_2, when $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$ we have that for each signal s:

$$\begin{aligned}R_s(n,\sigma) = &\pi_2^n\big(k_1(1-e_1) - k_2e_1\big)\Pr[\tau = r|s] \\ &-\Pr[\text{Piv}|\tau = a]\big((w-k_2)(1-e_2) - ce_2\big)\Pr[\tau = a|s] \\ &+\Pr[\text{Piv}|\tau = r]\big(c(1-e_1) - (w-k_2)e_1\big)\Pr[\tau = r|s]\end{aligned}$$

On the other hand under parameters w', c', k_1', k_2', e_1', e_2', $\varepsilon' = \varepsilon$, p_A', when $e_2' = 0, e_1' = 0$ and $k_2' = 0$ (the state of the world model analyzed in Proposition 4) we have that for each signal s:

$$R_s'(n,\sigma) = \pi_2^n k_1' \Pr'[\tau = r|s] - \Pr[\text{Piv}|\tau = a]w'\Pr'[\tau = a|s] + \Pr[\text{Piv}|\tau = r]c'\Pr'[\tau = r|s]$$

So let $k_1' = k_1(1-e_1) - k_2e_1$, $w' = (w-k_2)(1-e_2) - ce_2$, $c' = c(1-e_1) - (w-k_2)e_1$ and $p_A' = \frac{p_A - e_1}{(1-e_2)-e_1}$. Then $\Pr'[\tau = r|s] = \Pr[\tau = r|s]$ (and therefore also $\Pr'[\tau = a|s] = \Pr[\tau = a|s]$) , and therefore $R_s'(n,\sigma) = R_s(n,\sigma)$ for all n and σ.[18] Therefore, when $\frac{k_2}{k_1} = \frac{e_2}{1-e_2}$, Proposition 4 applies directly to w', c', p_A', k_1', p_A' and ε'.

Next, we show that the parameter set for which Proposition 4 applies is non-empty. Note that if the parameters w, c, k_1, ε, and p_A meet the two conditions of Proposition 4, then for sufficiently small (but positive) e_1 and e_2 and k_2, they will be met by w' c', p_A'. and k_1', p_A' and ε'. The reason is that as $e_1 \to 0$, $e_2 \to 0$ and $k_2 \to 0$, the parameters in the second problem converge to those of the first problem, and so do the corresponding boundaries of the open sets that define the two conditions.[19]

[18] $\pi_2^{\sigma^n}$, $\Pr[\text{Piv}|\tau = a]$ and $\Pr[\text{Piv}|\tau = r]$ are exactly the same in both expressions for all σ and n because $\varepsilon' = \varepsilon$.

[19] Concretely, pick w, c and k_1 which meet the conditions of proposition 3, given p_A and ε. Then set e_1, e_2 and k_2 low enough so that the conditions are met by w', c', k_1 and p_A'. If at this stage $\frac{k_2}{k_1}$ is smaller than $\frac{e_2}{1-e_2}$ then just decrease e_2 further in order to restore equality, and this can only make the conditions slacker. Similarly, if at this stage $\frac{k_2}{k_1}$ is bigger than $\frac{e_2}{1-e_2}$ then just decrease k_2 further in order to restore equality.

Applying Proposition 4 we therefore have that when $q = \frac{1}{2}$ truth-telling is an equilibrium of the second problem for all sufficiently large n (and therefore of the first one, as they are exactly the same). Furthermore, since $\varepsilon < \frac{1}{2} < 1 - \varepsilon$ we also have that in the limit $\pi_1 = 1$ and $\pi_2 = 0$. That is, the committee aggregates information perfectly in the sense that for sufficiently large n it approximates the state of the art with arbitrarily precision. ▮*(Lemma 6)*

References

Austen-Smith, David and Jeffrey Banks (1996), "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review*, 90, 34–45.

Brennan, Geoffrey and Philip Pettit (2004), *The Economy of Esteem*. An Essay on Civil and Political Society, Oxford University Press.

Callander, Steven (2008), "Majority rule when voters like to win." *Games and Economic Behavior*, 64, 393–420.

Ellingsen, Tore and Magnus Johannesson (2008), "Pride and Prejudice: The Human Side of Incentive Theory." *The American Economic Review*, 98, 990–1008.

Feddersen, Timothy and Wolfgang Pesendorfer (1997), "Voting Behavior and Information Aggregation in Elections With Private Information." *Econometrica*, 65, 1029–1058.

Feddersen, Timothy and Wolfgang Pesendorfer (1998), "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting." *The American Political Science Review*, 92, 23–35.

Gerling, Kerstin, Hans Peter Grüner, Alexandra Kiel, and Elisabeth Schulte (2005), "Information acquisition and decision making in committees: A survey." *European Journal of Political Economy*, 21, 563–597.

Huck, Steffen and Kai Konrad (2005), "Moral Cost, Commitment, and Committee Size." *Journal of Institutional and Theoretical Economics*, 161, 575–588.

Levy, Gilat (2007), "Decision Making in Committees: Transparency, Reputation, and Voting Rules." *The American Economic Review*, 97, 150–168.

Li, Hao (2001), "A Theory of Conservatism." *Journal of Political Economy*, 109, 617–636.

Li, Hao and Wing Suen (2009), "Viewpoint: Decision-making in committees." *Canadian Journal of Economics/Revue canadienne d'économique*, 42, 359–392.

McLennan, Andrew (1998), "Consequences of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents." *The American Political Science Review*, 92, 413–418.

Midjord, Rune, Tomás Rodríguez Barraquer, and Justin Valasek (2014), "Over-Caution of Large Committees of Experts." *CESifo Working Paper No. 4810*.

Morgan, John and Felix Várdy (2012), "Mixed Motives and the Optimal Size of Voting Bodies." *Journal of Political Economy*, 120, 986–1026.

Ottaviani, Marco and Peter Sorensen (2001), "Information aggregation in debate: who should speak first?" *Journal of Public Economics*, 81, 393–421.

Palfrey, Thomas and Howard Rosenthal (1985), "Voter Participation and Strategic Uncertainty." *The American Political Science Review*, 79, 62–78.

Peleg, Bezalel and Shmuel Zamir (2012), "Extending the condorcet jury theorem to a general dependent jury." *Social Choice and Welfare*, 39, 91–125.

Visser, Bauke and Otto H Swank (2007), "On Committees of Experts." *The Quarterly Journal of Economics*, 122, 337–372.