Bonn Econon Discussion Papers

Discussion Paper 02/2013

Consistency and Communication in Committees

by

Inga Deimen, Felix Ketelaar, Mark T. Le Quement

February 2013

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Kaiserstrasse 1
D-53113 Bonn

Financial support by the
Deutsche Forschungsgemeinschaft (DFG)
through the
Bonn Graduate School of Economics (BGSE)
is gratefully acknowledged.

Deutsche Post World Net is a sponsor of the BGSE.

Consistency and Communication in Committees*

Inga Deimen[†] Felix Ketelaar[‡] Mark T. Le Quement[§]

February 6, 2013

Abstract

This paper analyzes truthtelling incentives in pre-vote communication in heterogeneous committees. We generalize the classical Condorcet jury model by introducing a new informational structure that captures consistency of information. In contrast to the impossibility result shown by Coughlan (2000) for the classical model, full pooling of information followed by sincere voting is an equilibrium outcome of our model for a large set of parameter values implying the possibility of ex post conflict between committee members. Furthermore, abandonning the assumption of sincere voting, we characterize necessary and sufficient conditions for the implementability of the first best decision rule via truthful equilibria.

Keywords: Communication, Committees, Voting

JEL Classification: D72, D82, D83

*We thank in particular D. Szalay for his insights and encouragement. We are grateful to P. Esö, L. Felli, D. Krähmer, M. Meyer-Ter-Vehn for helpful comments and suggestions. This paper has benefitted from comments made by the following audiences: SFB meeting Spring 2012, Spring Meeting of Young Economists 2012, Bonn Micro Workshop, 11th meeting of the Society of Social Choice and Welfare, EEA-ESEM Congress 2012, Verein für Socialpolitik Jahrestagung 2012, TU Dortmund Volkswirtschaftliches Kolloquium, EDP Jamboree 2012.

†Bonn Graduate School of Economics, University of Bonn, Kaiserstraße 1, 53113 Bonn, Germany. E-Mail: ideimen@uni-bonn.de.

‡Bonn Graduate School of Economics, University of Bonn, Kaiserstraße 1, 53113 Bonn, Germany. E-Mail: 3fekete@uni-bonn.de.

§Institute of Microeconomics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany. E-Mail: mlequem@uni-bonn.de.

1 Introduction

Many situations of decision making are characterized by an asymmetry between a complex decision environment and simple decisions, such as making a particular investment or not, choosing between job candidates, and acquitting or convicting a defendant. In such scenarios, multiple contingencies, rich information and sophisticated language contrast with a limited, often binary action space. We show that the combination of fine information and coarse decisions can have important implications for information aggregation in heterogeneous committees: it mitigates conflict in the sense of enabling individuals with diverging preferences to truthfully share their information.

A key feature of rich information in a complex environment is its consistency: aggregate information in favor of a particular action will be more or less convincing depending on the extent to which individual pieces of information draw a coherent picture. We incorporate this aspect into a model of collective decision making in which privately informed heterogeneous agents communicate through cheap talk before casting their vote. Our main finding is that in our model, heterogeneity and ex post conflict are frequently compatible with full information pooling. Our possibility results contrast with the negative insights provided by existing literature for coarser environments (Coughlan (2000), Austen-Smith and Feddersen (2006), Meirowitz (2007), Van Weelden (2008)).

Consider the following scenario. A jury aiming at determining whether a defendant is guilty or not seeks to establish a set of key facts about the crime. Among others, these include its location in time and space. Along each of these two dimensions, a plurality of options may be considered. Was the crime committed in the morning or in the evening of a particular day, on Monday or Tuesday of a given week? Did it happen outside or inside a given house, in room A only, in B only or in both? Conditional on the defendant being guilty, for each of these dimensions only one of all the possible answers can be true while all others must be false. In other words, different locations of the crime in time and space constitute mutually exclusive modalities of the guilty state. Other relevant dimensions of the crime will be subject to the same insight. In general, each of the basic states of the classical Condorcet jury model (guilt or innocence) can be regarded as a set of mutually exclusive substates, each substate representing a separate instance of the basic state that it incarnates.

In our model, evidence pointing towards the same basic state of the world exhibits varying degrees of coherence. The more consistently signals indicate a given substate, the stronger the evidence for the basic state that it is an instance of. From a payoff perspective, however, jurors do not as such care about which modality of guilt or innocence applies. They simply wish to establish with sufficient certainty whether the defendant is guilty or not. In other words, the substates constituted by the different modalities of each basic state are payoff irrelevant.

We provide three main types of results. Theorem 1 provides necessary and sufficient conditions for the existence of the truthful communication and sincere voting equilibrium (TS equilibrium) in the presence of a positive ex ante probability of ex post conflict among jurors. Within our setup, we find that the TS equilibrium exists for a large set of parameter values. Theorem 2, in contrast to existing results in the literature (Coughlan (2000), Le Quement (2012)), shows that the number of aggregate signal profiles at which conflict arises at the voting stage is an imperfect indicator of the difficulty of achieving full information pooling in a heterogeneous committee. In addition, we identify conditions under which our result implies that increasing committee size ultimately guarantees the existence of the TS equilibrium, for a fixed constellation of preference types. In Theorems 3 and 4, we abandon the assumption of sincere voting and consider coordinated voting equilibria featuring weakly dominated voting. Theorem 4 identifies the necessary and sufficient conditions under which the welfare maximizing decision rule can be implemented. We furthermore point out that if this rule is implementable by any mechanism, then it is implementable in a truthful equilibrium of the simple communication and voting game. The identified implementability conditions are independent of the chosen voting rule (excluding unanimous rules) and are satisfied for a large set of parameter values.

Our analysis uncovers new strategic effects. In the putative equilibrium of the classical model, the pivotality of a given juror in the communication stage pins down uniquely the information held by remaining committee members. In our model, this uniqueness breaks down because multiple signal profiles generate comparable posterior probabilities of guilt. This follows from the impact of consistency on Bayesian updating within our model. The multiplicity of pivotal scenarios allows two forces to arise, each of which incentivizes truthtelling. First, for a subset of pivotal signal profiles, all jurors may agree with the decision following from a truthful announcement, thus not wishing to deviate at these profiles. We call this the consensus effect. Secondly, among the set of pivotal signal profiles faced by a juror, a given deviation from truthtelling may overturn a conviction at one subset and overturn an acquittal at another. A given deviation will, in other terms, not necessarily have a systematic impact on the outcome. This unpredictability generates a risk to deviating. We call this the uncertainty effect. In contrast, the two effects described above do not arise in the classical setup. There, at the unique pivotal signal profile, at least one juror type will always disagree with the decision following from fully shared information and sincere voting. Furthermore, the effect of any announcement on the likelihood of conviction is perfectly predictable. Accordingly, in the putative TS equilibrium of the classical model, at least one juror type will deviate from truthtelling, thereby increasing the likelihood that his favorite decision is taken.

From a methodological perspective, our model can be interpreted as a generalization of the classical model. We show in Section 7 that the classical model can be nested in a generalized version of ours. In fact, abandoning either the rich signal structure or the rich message space leads back to the impossibility result of Coughlan (2000). In this sense, our contribution essentially resides in the identification of new effects arising from the interplay of the complexity of the decision environment (contingencies, information, language) and the coarseness of the action space.

We proceed as follows. Section 1 closes with a literature review. Section 2 presents the model. Section 3 develops a simple formal example identifying key mechanisms. Section 4 analyzes the role of consistency in Bayesian updating. Section 5 presents our main results on the existence of the TS equilibrium. Section 6 relaxes the assumption of sincere voting and considers the possibility to implement the first best decision rule through truthful equilibria. Section 7 explicitly relates our model to the classical Condorcet setup. Section 8 concludes. All proofs are relegated to the Appendix.

Related literature. Building on the theory of strategic voting as information aggregation (Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1998)), a milestone in the literature on cheap talk deliberation and collective decisions in heterogeneous committees is the impossibility result presented in Coughlan (2000). The latter states that in the classical binary collective decision problem, full information sharing followed by sincere voting by committee members cannot be an equilibrium outcome if jurors do not agree on the optimal decision for all profiles of pooled information. A strictly positive ex ante probability of ex post conflict between jurors, even arbitrarily low, suffices to cause a breakdown of the TS equilibrium.

At an abstract level, our paper belongs to a class of contributions that modify the classical model and reestablish the TS equilibrium prediction under heterogeneous preferences. While our paper examines the role of informational consistency, Austen-Smith and Feddersen (2006) add the realistic feature of uncertainty about the preferences of jury members. They show that this can render full pooling combined with sincere voting possible, as long as the voting rule is not unanimity. In a complementary contribution, Meirowitz (2007) shows that the TS equilibrium exists if individual jurors are sufficiently confident that the majority of jurors shares their own preferences. Van Weelden (2008) however adds an important caveat: when communication is sequential, uncertainty does not anymore suffice to ensure the existence of the TS equilibrium. In our environment, as we point out in Section 3, sequential communication does not always eliminate the TS equilibrium. Furthermore, Le Quement (2012) points out a second caveat to Austen-Smith and Feddersen (2006), showing that only minimal disagreement is compatible with the TS equilibrium in large heterogeneous committees. This latter caveat does not apply to our model, where an unbounded number of conflictual aggregate signal profiles is compatible with the TS equilibrium in sufficiently large committees.

Another class of contributions approaches the communication problem from a mechanism design perspective. In Gerardi and Yariv (2007), a mediator centralizes the private reports of potentially heterogeneous jurors and subsequently recommends an identical voting decision to all jurors in the final voting stage. Using such a mediator, information is thus aggregated before the vote independently of the voting rule (except for unanimity). In line with this result, in our analysis of truthful equilibria featuring weakly dominated voting, we indeed find that all non-unanimous voting rules are equivalent. In Wolinsky (2002), truthtelling requires the implementation of an ex post inefficient decision rule which generates pivotal scenarios in which lying is costly. In our environment, already the ex post efficient decision rule will typically generate pivotal scenarios incentivizing truthtelling. In Gerardi, McLean and Postlewaite (2009), a mediator uses the correlation among signals to threaten heterogeneous individuals with punishment if their report does not match other experts' report. In contrast to this sophisticated protocol, the optimal mechanism in our setup takes the simple form of a truthful equilibrium of the communication and voting game.

A third class of contributions maintains the classical model but examines a different communication scenario. In Hummel (2012) as well as Le Quement and Yokeeswaran (2012), a heterogeneous committee splits up into subgroups of homogeneous members in the deliberation phase, thus achieving local sharing of information. Given our positive results on the existence of welfare optimizing equilibria that involve full pooling of information, we omit the analysis of partial pooling scenarios.

A set of positive and normative contributions stresses the importance of the full pooling scenario. The experimental work of Goeree and Yariv (2011) documents extensive truthtelling in heterogeneous committees and finds that individuals assign substantial weight to the information revealed by others. Dickson, Hafer and Landa (2008) similarly find evidence of intense sharing of information among heterogeneous jurors. The philosophical literature on deliberation (e.g. Habermas (1990), Elster (1997), Manin (1987)) assigns an intrinsic value to exhaustive deliberations conducive to full exchange of information.

Finally, we mention two contributions on cheap talk communication that bear some relation to our work. Battaglini (2003) considers a setup in which multidimensional state, signal and action spaces generically allow for full information extraction from multiple experts, contrasting the one-dimensional baseline model of Crawford and Sobel (1982). While the insight that a richer informational environment facilitates truthful revelation is reminiscent of our results, the underlying mechanisms as well as several structural assumptions are fundamentally different. For instance, Battaglini's model does not exhibit an asymmetry between fine information and a coarse action space. Furthermore, in our model, there is no sense in which preferences and biases are multidimensional. In Blume, Board and Kawamura (2007), the information transmission process is subject to noise. The authors show that this may improve the ex ante payoffs of sender and receiver by beneficially affecting truthtelling incentives. Our paper shares with the latter the insight that uncertainty about the impact of messages may encourage truthtelling. However, while this uncertainty is generated by exogenously imposed noise in the above paper, it arises endogenously in our model: an individual message is interpreted in the light of other messages.

2 The model

We present our model using the notions of a standard jury trial setup. A jury of size $n \in \mathbb{N}$ is asked to decide whether to acquit (A) or convict (C) a defendant. The defendant is either innocent (I) or guilty according to modality 1 (G_1) or modality 2 (G_2) with prior probability $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. We denote by $\Omega = \{I, G_1, G_2\}$ the set of states of the world with typical element ω. For $\omega \in \{G_1, G_2\}$ we simply say the defendant is guilty. Our stylized assumptions on the state space and the prior are made for the purpose of simplicity and suffice to generate the main effects that we wish to study.

The jury decides about which action $a \in \{A, C\}$ to implement according to some prespecified voting rule $k \in \{2, \ldots, n-1\}$. Each juror $j \in \{1, \ldots, n\}$ casts a vote in favor of one of the two actions. If the number of votes cast for conviction is larger or equal to k, the defendant is convicted while otherwise he is acquitted. We rule out the two types of unanimity voting rules and discuss the reasons for this omission after Theorem 1.

The utility of juror j depends on the underlying state, the implemented action and an individual preference parameter $q_j \in (0,1)$ in the following way:

$$u_j(a,\omega) = \begin{cases} 0 & \text{for } (a,\omega) \in \{(A,I),(C,G_1),(C,G_2)\}, \\ -q_j & \text{for } (a,\omega) = (C,I), \\ -(1-q_j) & \text{for } (a,\omega) \in \{(A,G_1),(A,G_2)\}. \end{cases}$$

While utilities of correct decisions (acquitting an innocent resp. convicting a guilty defendant) are normalized to 0, the relative loss from making a mistake by convicting an innocent resp. acquitting a guilty defendant is determined by the preference parameter q_j. As juror j maximizes expected utility, he prefers conviction over acquittal if and only if the probability of the defendant being guilty exceeds q_j. The parameter q_j can hence be interpreted as a "threshold of reasonable doubt". Note in particular that, for each possible action, juror j is indifferent as to whether the defendant is guilty according to modality 1 (G_1) or modality 2 (G_2).

For ease of presentation, we assume that the jury consists of only two preference types, doves and hawks, whose respective preference parameters are given by $q_D \in (0,1)$ and $q_H \in (0,1)$. Doves are assumed to require higher evidence for guilt than hawks to prefer conviction over acquittal, i.e. $q_D > q_H$. In Section 6 we extend our results to the case of individual preference parameters.

Prior to the voting stage, each juror j receives a private signal from $\{i, g_1, g_2\}$. Signals are i.i.d. conditional on the realized state of the world ω. An individual signal indicates the correct state of the world with probability $p \in \left(\frac{1}{3}, 1\right)$ and indicates either of the remaining states with probability $\frac{(1-p)}{2}$. We refer to Section 7 for a discussion of our assumptions on the signal generating process.

The sum of all jurors' individual signals constitutes a *signal profile* (x, y, z) with $x + y + z = n$, where x denotes the total number of i-signals, y the total number of g_1-signals and z the total number of g_2-signals held by the committee. In particular, the Bayesian posterior probability assigned to the defendant being guilty (i.e. $\omega \in \{G_1, G_2\}$) is given as

$$\beta(x, y, z) = \frac{\left(\frac{2p}{1-p}\right)^y + \left(\frac{2p}{1-p}\right)^z}{\left(\frac{2p}{1-p}\right)^x + \left(\frac{2p}{1-p}\right)^y + \left(\frac{2p}{1-p}\right)^z}.$$

As jurors do not differentiate between state G_1 and state G_2 in terms of utilities, the number $\beta(x, y, z)$ is a sufficient statistic for the preferred action of each individual juror.

The difference between the two numbers y and z captures the notion of consistency of signal profiles and plays a key role in our model. Section 4 offers an in-depth analysis of the properties of our information structure. In Section 7, we relate our environment to a setting where information about the modality of guilt is garbled.

After having received their signals, jurors communicate through a round of simultaneous and public cheap talk. Given our focus on truthful equilibria, it is without loss of generality to assume that each juror j sends a message from $\{i, g_1, g_2\}$.
To summarize, the timing of the game is as follows:

0. Nature draws a state of the world.
1. Each juror receives a private signal.
2. Each juror simultaneously emits a public cheap talk message.
3. Each juror casts a vote.
4. An action is implemented according to the voting rule.

Our equilibrium concept is Perfect Bayesian Equilibrium. The core of our analysis is concerned with the existence of the following particular equilibrium: jurors truthfully reveal their private information at the communication stage, jurors (correctly) believe that other jurors have revealed their private information truthfully, and juror j votes sincerely, i.e. votes for conviction if and only if the probability of guilt of the defendant exceeds q_j, given his private information and the reports of other jurors. We call this putative equilibrium the TS equilibrium. Section 6 drops the assumption of sincere voting.

3 A simple example

In this section, we present a simple example that demonstrates the key forces at work in our model and in particular highlights the two potential sources of truthtelling described in the introduction: the consensus and uncertainty effects.

Consider a three persons jury consisting of one hawk and two doves. The voting rule is given by simple majority, i.e. $k = 2$. Aggregate signal profiles can be ordered exhaustively with respect to the conditional probability of guilt that they induce, i.e.

$$\beta(3,0,0) < \begin{matrix} \beta(2,0,1) \\ \beta(2,1,0) \end{matrix} < \beta(1,1,1) < \begin{matrix} \beta(1,0,2) \\ \beta(1,2,0) \end{matrix} < \begin{matrix} \beta(0,1,2) \\ \beta(0,2,1) \end{matrix} < \begin{matrix} \beta(0,0,3) \\ \beta(0,3,0) \end{matrix}.$$

Preference parameters q_H, q_D are assumed such that a dove favors conviction if and only if the aggregate signal profile is either $(0,3,0)$ or $(0,0,3)$ while a hawk favors conviction if and only if the aggregate signal profile is $(0,3,0), (0,0,3), (0,1,2)$ or $(0,2,1)$. There are thus two signal profiles for which hawks and doves disagree on the optimal decision, namely $(0,1,2)$ and $(0,2,1)$.

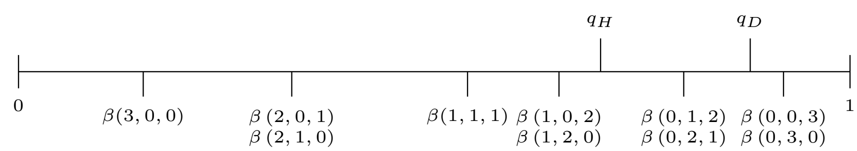

Figure 1: Conditional probabilities of guilt.

In this setting, TS strategies and beliefs always constitute an equilibrium, for any parameter values q_H and q_D consistent with the above preferences and for any signal precision $p > \frac{1}{3}$. We prove this by verifying explicitly that in a putative TS equilibrium, no individual juror has an incentive to deviate from truthtelling followed by sincere voting.

Note the following three simple facts. First, whatever the announcement made by a juror in the communication stage (truthful or not), he has no incentive to subsequently deviate from sincere voting, as such a deviation decreases the probability that his favored decision ensues. In other words, deviating at the voting stage is a weakly dominated strategy. Secondly, given that the voting rule is simple majority, doves are always able to implement their favored decision. Indeed, if the hawk favors conviction but doves do not, the latter can implement acquittal simply by jointly voting for it while if the hawk prefers acquittal, doves do so as well. Hence the doves have no incentive to deviate from truthtelling. Finally, any rational juror chooses his action conditional on being pivotal.

We now analyze the truthtelling incentives of the hawk in the putative TS equilibrium. The hawk's announcement is pivotal if the remaining two jurors hold signal profiles $(0,2,0)$ or $(0,0,2)$. In the first (second) case, a g_1- (g_2-) announcement would cause conviction while any of the remaining announcements would cause acquittal. To systematically analyze the hawk's incentive to deviate at each of his three possible information sets i, g_1 and g_2, note that given the symmetry of the model, conditions ensuring truthtelling of the hawk holding a g_1-signal or a g_2-signal are identical modulo an exchange of subscripts. We can therefore restrict our analysis to deviations of the hawk holding an i-signal or holding a g_1-signal.

Assume that the hawk holds an i-signal and is pivotal at the communication stage. The signal profile of the entire committee is then $(1, 2, 0)$ or $(1, 0, 2)$. In either case, the decision that is taken by the committee given the true signal profile is acquittal and coincides with the decision favored by the hawk. Accordingly, he has no incentive to deviate from truthtelling. Here, the consensus effect is the sole source of truthtelling: despite heterogeneity and despite the existence of signal profiles generating conflict, a hawk juror with an i-signal fully agrees with the doves on the preferred action in all pivotal scenarios.

Assume next that the hawk holds a g_1-signal and is pivotal in the communication stage. The signal profile of the entire committee is then either $(0, 1, 2)$ or $(0, 3, 0)$. Here, in contrast to the previous case, the hawk disagrees with the acquittal decision ensuing from truthtelling at $(0, 1, 2)$ while he agrees with the conviction decision ensuing from truthtelling at $(0, 3, 0)$. If the hawk deviates to announcing an i-signal, the signal profile observed at the voting stage by other jurors is either $(1, 0, 2)$ or $(1, 2, 0)$, thus leading to an undesired acquittal. The deviation to an i-report is therefore dominated by truthtelling. If the hawk deviates to a g_2-announcement, the signal profile observed at the voting stage by the remaining jurors is given by $(0, 0, 3)$ or $(0, 2, 1)$. The deviation beneficially overturns an acquittal in the first case but adversely overturns a conviction in the second case. The hawk thus faces uncertainty about the impact of his statement: while at pivotal profile $(0, 0, 2)$, a g_2-report is harsher than a g_1-report and constitutes the only way to induce the desired conviction, the situation is exactly reversed at pivotal profile $(0, 2, 0)$.

Among the two pivotal profiles $(0, 0, 2)$ and $(0, 2, 0)$ faced by the hawk when holding a g_1-signal, $(0, 0, 2)$ thus incentivizes lying while $(0, 2, 0)$ incentivizes truthtelling. Which incentive dominates depends on the relative probability assigned to these two profiles, the latter itself depending on the probability assigned to the states G_1 and G_2. A juror holding a g_1-signal assigns a higher probability to state G_1 than to state G_2, and accordingly to profile $(0, 2, 0)$ than to profile $(0, 0, 2)$. The signal profile incentivizing truthtelling is thus assigned a higher probability than the one incentivizing lying. Hence the hawk, when holding a g_1-signal, never prefers to announce a g_2-signal. We may conclude that the TS equilibrium exists for all parameter values matching our assumptions, despite the existence of signal profiles generating conflict.

We close the discussion of this example with a remark on sequential communication. The prescribed equilibrium continues to exist under a sequential communication protocol where the single hawk speaks first. Indeed, the hawk's incentives when speaking first are identical to those under the simultaneous protocol while the doves still determine the outcome and hence have no incentives to deviate. This insight stands in contrast to the impossibility result of van Weelden (2008) for the classical setup with unknown preference types.

4 Properties of the information structure

In this section, we analyze the properties of conditional probabilities arising within our model. The features identified in this section are essential to the subsequent equilibrium analysis. In a first step, we formalize the notion of consistency and provide a qualitative result regarding its impact on ex post probabilities in committees of fixed size. The presence of a consistency concern is the central force behind our results. Indeed, eliminating the consistency concern from our model leads back to a setting in which Coughlan's impossibility result applies, as we point out in Section 7. In a second step, we relax the assumption of fixed committee size and consider the relative probabilities of the states of the world as the only relevant quantities. This leads to a remarkably simple lexicographic ordering of signal profiles. This ordering lies at the heart of our asymptotic existence results.

Recall that we denote a signal profile by (x, y, z), where the entries describe respectively the numbers of i-, g_1- and g_2-signals. The Bayesian posterior probability of guilt, for a signal profile (x, y, z), is given by

$$\beta(x, y, z) = \frac{\left(\frac{2p}{1-p}\right)^y + \left(\frac{2p}{1-p}\right)^z}{\left(\frac{2p}{1-p}\right)^x + \left(\frac{2p}{1-p}\right)^y + \left(\frac{2p}{1-p}\right)^z}.$$

Clearly, as $p > \frac{1}{3}$ and hence $\frac{2p}{1-p} > 1$, $\beta(x, y, z)$ is decreasing in x and increasing in y and z. This captures the straightforward intuition that a larger number of i-signals decreases the probability of guilt while a larger number of either g-signal increases the probability of guilt.

For a given committee, the number of jurors n and hence the total number of signals $x+y+z=n$ is fixed. As we are ultimately interested in the reporting incentives of a juror, we now analyze the effect of shifting mass from one entry of β to another. Indeed, subtracting one unit from a given entry of β and adding it to another replicates the change in beliefs of other jurors achievable by an individual juror misreporting his signal in the TS equilibrium.

It immediately follows from the monotonicity properties of β that shifting mass from the y- or z-entry to the x-entry of β (or vice versa) decreases (increases) the posterior probability of guilt. In other words, misreporting a guilty signal as an innocent signal (and vice versa) in the putative TS equilibrium decreases (increases) the posterior beliefs of other jurors regarding the probability of guilt and thus unilaterally increases the likelihood of acquittal (conviction). Considering now the effect of shifting mass between the last two entries, note that this shift keeps the total number of guilty signals $n_g(x, y, z) := y + z$ constant. Such a shift solely affects the *consistency* $\Delta(x, y, z) := |z - y|$ of signals indicating guilt. Rewriting $\beta(x, y, z)$ as

$$\beta(n, n_g, \Delta) = \frac{\left(\frac{2p}{1-p}\right)^{\frac{n_g+\Delta}{2}} + \left(\frac{2p}{1-p}\right)^{\frac{n_g-\Delta}{2}}}{\left(\frac{2p}{1-p}\right)^{n-n_g} + \left(\frac{2p}{1-p}\right)^{\frac{n_g+\Delta}{2}} + \left(\frac{2p}{1-p}\right)^{\frac{n_g-\Delta}{2}}},$$

the following lemma allows us to evaluate the effect of a shift of mass between guilty signals.

Lemma 1. *For fixed n and Δ, $\beta = \beta(n, n_g, \Delta)$ is increasing in n_g. For fixed n and n_g, $\beta = \beta(n, n_g, \Delta)$ is increasing in Δ.*

Lemma 1 clearly points out that two separate forces drive the posterior probability of guilt. In particular, when comparing two signal profiles, a larger total number n_g of g-signals may be offset by a lack of consistency Δ among g-signals. While a direct quantitative comparison of these two forces is possible, we introduce in what follows a simpler method for the comparison of ex post probabilities of guilt arising from different signal profiles. Define

$$\delta_1(x, y, z) := \max\{y, z\} - x,$$
$$\delta_2(x, y, z) := \min\{y, z\} - x$$

and note that

$$\beta(x, y, z) = \beta(\delta_1, \delta_2) = \frac{\left(\frac{2p}{1-p}\right)^{\delta_1} + \left(\frac{2p}{1-p}\right)^{\delta_2}}{1 + \left(\frac{2p}{1-p}\right)^{\delta_1} + \left(\frac{2p}{1-p}\right)^{\delta_2}}.$$

By construction, δ_1 measures the relative likelihood of the ex post more likely modality of guilt compared to innocence. Similarly, δ_2 measures the relative likelihood of the ex post less likely modality of guilt compared to innocence. Under a moderate assumption on signal precisions, this leads to a strikingly simple ordering of signal profiles of possibly different cardinalities:

Lemma 2. *Assume $p \geq \frac{1}{2}$ and let $(x, y, z), (\tilde{x}, \tilde{y}, \tilde{z}) \in \mathbb{N}^3$. Then*

$$\beta(x, y, z) > \beta(\tilde{x}, \tilde{y}, \tilde{z}) \Leftrightarrow \begin{cases} \delta_1(x, y, z) > \delta_1(\tilde{x}, \tilde{y}, \tilde{z}) \quad \textit{or} \\ \delta_1(x, y, z) = \delta_1(\tilde{x}, \tilde{y}, \tilde{z}) \wedge \delta_2(x, y, z) > \delta_2(\tilde{x}, \tilde{y}, \tilde{z}). \end{cases}$$

For reasonable information quality, the lexicographic structure implied by Lemma 2 sheds light on the heuristic followed by a juror: first, single out the most likely modality of guilt. Next, compare its likelihood to the likelihood of innocence. Only then, if necessary, consider the less likely modality of guilt and assess its relative probability.

As a consequence of the lexicographic structure of the signal space with respect to δ_1 and δ_2 and in stark contrast to the classical Condorcet model, the following asymptotic feature arises in our model: for large n, limit points of ex post probabilities of guilt appear strictly within the $[0, 1]$-interval of ex post probabilities. Indeed, as the committee size n increases, there exist arbitrarily many signal profiles characterized by the same $\delta_1 \in \mathbb{Z}$, and $\delta_2 \in \mathbb{Z}$ can become arbitrarily small, for any given δ_1.

Lemma 3. *Let $p \geq \frac{1}{2}$ and let $\beta(x, y, z) = \beta(\delta_1, \delta_2)$ denote the ex post probability of guilt arising from signal profile $(x, y, z) \in \mathbb{N}^3$. Then, the set of ex post probabilities has limit points precisely at $0, 1$ and at $\lim_{\delta_2 \to -\infty} \beta(\delta_1, \delta_2) = \frac{\left(\frac{2p}{1-p}\right)^{\delta_1}}{1+\left(\frac{2p}{1-p}\right)^{\delta_1}}$, for any $\delta_1 \in \mathbb{Z}$.*

In Coughlan (2000), for a given q_D and q_H, as n tends to infinity, the number of signal profiles generating ex post conflict remains fixed. In contrast, in our model, the number of such profiles will typically increase as n increases. Any additional profile generating ex post conflict has a negative influence on truthful reporting incentives. Nonetheless, as we show in Theorem 2 in the next section, the detrimental effect of additional profiles generating ex post conflict is frequently outbalanced by the appearance of new profiles that incentivize truthtelling. In fact, a sufficiently large committee size may even serve as a sufficient condition to guarantee the existence of a TS equilibrium, as we point out in the discussion of Theorem 2.

5 Analysis of the TS equilibrium

The example of Section 3 shows that the TS equilibrium can exist despite potential disagreement after full pooling of information. Making use of the general properties identified in Section 4, this section provides an equilibrium analysis for arbitrary committee sizes. Section 5.1 introduces key notions, Section 5.2 establishes our main results concerning the existence of the TS equilibrium.

5.1 Key notions

Fix the number n of jurors. For any preference type q_j, $j \in \{H, D\}$, unless it prefers acquittal for any possible signal realization, there exists a *threshold profile* (x_j, y_j, z_j) such that the following holds: a juror of preference type q_j prefers acquittal for signal profile (x, y, z) if $\beta(x, y, z) < \beta(x_j, y_j, z_j)$ while he prefers conviction if $\beta(x, y, z) \geq \beta(x_j, y_j, z_j)$. That is, a juror of preference type q_j prefers conviction precisely for those signal profiles that yield at least as much evidence for the defendant being guilty as his threshold profile (x_j, y_j, z_j) does. Threshold profiles are unique up to transposition of the last two entries. Without loss of generality, we focus on threshold profiles satisfying $z_j \geq y_j$.

We say that a signal profile (x, y, z) is a *conflict profile* if conditional on signal profile (x, y, z) hawks and doves disagree on the preferred action, that is, if

$$q_H \leq \beta(x, y, z) < q_D \quad \text{resp.} \quad \beta(x_H, y_H, z_H) \leq \beta(x, y, z) < \beta(x_D, y_D, z_D).$$

The number of conflict profiles provides a measure of conflict within the committee that abstracts from the numerical values of q_H and q_D but directly relates to the informational setup of the model. Recall that in the classical model, the existence of the TS equilibrium is compatible with some degree of heterogeneity among jurors' preference parameters q_H, q_D but incompatible with the existence of a conflict profile.

Given the assumption that there are only two types of jurors, the voting rule k, as long as it is not unanimous, matters only in a binary sense: either the number of hawks matches or exceeds the number of votes k required for conviction, so that hawks are sufficiently numerous to implement conviction whenever they wish. Otherwise, if hawks are not sufficiently numerous, the favored decision of the doves is always implemented as they can veto any undesired attempt from the hawks to convict the defendant. We say that hawks have *critical mass* in the first scenario while doves have critical mass in the second scenario. Note that in a putative TS equilibrium, the outcome of the trial will be as follows: if the group having critical mass decides according to threshold profile (x_j, y_j, z_j), the defendant will be convicted if and only if the revealed signal profile (x, y, z) satisfies $\beta(x_j, y_j, z_j) \leq \beta(x, y, z)$.

5.2 Existence

This section provides necessary and sufficient conditions for the existence of the TS equilibrium. We find that the TS outcome frequently constitutes an equilibrium of our model and is compatible with an arbitrary number of conflict profiles. For the purpose of examining truthtelling incentives, an exhaustive characterization of the set of pivotal profiles is required. Lemma A in the Appendix provides an explicit classification of all pivotal profiles.

Verifying the existence of the TS equilibrium essentially consists in analyzing one of the following two cases. If hawks have critical mass, the reporting incentives of a dove holding a g_1- or a g_2-signal must be examined. If doves have critical mass, the reporting incentives of a hawk holding an i-signal must be examined. The involved deviation scenarios are intuitive; they correspond to a juror's incentive to bend the jury's decision in the direction of his own relative bias. Note furthermore that these deviation scenarios are analogues of those determining a breakdown of the TS equilibrium within the classical model. We provide a treatment of these deviation incentives in the proofs of our theorems.

In the steps preceding Theorem 1, we rule out all further deviations in the putative TS equilibrium. First, by definition no juror has an incentive to deviate at the voting stage. Secondly, no juror of the preference type that has critical mass has an incentive to deviate at the communication stage. Indeed, the outcome in a putative TS equilibrium always coincides with the preferred outcome of the group having critical mass. Thirdly, the following lemma rules out deviations across guilty signals.

Lemma 4. *In the putative TS equilibrium, a juror holding a g_2-signal (g_1-signal) never has an incentive to deviate by reporting a g_1-signal (g_2-signal).*

The argument behind Lemma 4 relies on the uncertainty effect described in the example of Section 3. A single juror is uncertain about the relative impact of either g-report. If, among other jurors' reports, g_i-signals are predominant as compared to g_j-signals, then a g_i-announcement will be more effective at triggering a conviction than a g_j-announcement. A juror holding a g_i-signal assigns a higher probability to state G_i than to G_j. He accordingly assigns a higher probability to g_i-announcements being predominant as compared to g_j-announcements, and thus finds it optimal to report a g_i-signal in order to maximize the likelihood of a conviction conditional on the defendant being guilty. Summing up, the key insight is that a juror holding a g_i-signal assigns higher probability to scenarios incentivizing truthtelling than to those incentivizing misreporting a g_j-signal.

Given Lemma 4, relevant deviations can either consist in reporting an i-signal instead of some g-signal, which will unilaterally increase the chance of an acquittal, or they can consist in reporting some g-signal instead of an i-signal, which will unilaterally increase the chance of a conviction. Hence, in the putative TS equilibrium, neither a dove holding an i-signal nor a hawk holding a g_1- or a g_2-signal has an incentive to deviate: hawks (doves) never wish to overturn a conviction (an acquittal) by deviating from truthtelling.

We now present our first main result.

Theorem 1. *Let hawks have critical mass.*

a) For any hawk type q_H the TS equilibrium exists if and only if the value of q_D lies below a given upper bound $\hat{q}_D(q_H) > q_H$.

b) For any q_H s.t. $(x_H, y_H, z_H) \notin \{(0,0,n), (n-1,0,1)\}$, the pair (q_H, q_D) with $q_D = \hat{q}_D(q_H)$ implies at least one conflict profile.

Theorem 1 provides a general existence result for the TS equilibrium. Part a) states the existence of a critical dove type $\hat{q}_D(q_H)$ such that the TS equilibrium exists if and only if $q_D \in (q_H, \hat{q}_D(q_H)]$. An analytical expression for this critical dove type is given in the proof. Part b) yields a fundamental qualitative statement: for any hawk type $q_H \in (\beta(n-1,0,1), \beta(0,1,n-1))$, there exist dove types q_D such that the TS equilibrium exists despite the fact that they imply at least one conflict profile. Note that for $n \to \infty$ we have $(\beta(n-1,0,1), \beta(0,1,n-1)) \to (0,1)$. Part b) of Theorem 1 stands in stark contrast to Coughlan's impossibility result in demonstrating that in our environment, the TS equilibrium is generically compatible with the existence of at least one conflict profile.

The proof of Theorem 1 relies on the consensus effect by excluding a dove's deviation from some g-signal to an i-signal. Regarding Part a), note that a dove's utility from conviction is linear and strictly decreasing in q_D for any signal profile (x, y, z) and recall that a switch from some g-report to an i-report unilaterally decreases the likelihood of a conviction. Furthermore, for the limit case of perfectly aligned preferences, i.e. $q_D = q_H$, truthtelling is obviously optimal as it ensures the preferred decision for the doves as well as for the hawks at any signal profile.

As for Part b), consider the case where $q_D = \beta(x_H, y_H, z_H) > q_H$. Here, doves still weakly prefer the hawks' favorite outcome for any signal profile, with indifference precisely for signal profiles (x_H, y_H, z_H) and (x_H, z_H, y_H). A marginal increase of q_D therefore yields a marginal incentive to change the outcome from conviction to acquittal for these two conflict profiles by deviating to an i-report. However, by the assumptions on the hawks' threshold profile (x_H, y_H, z_H), there exists another signal profile for which a deviation towards an i-report changes the outcome (see Lemma A in the Appendix). This change yields a strict incentive for truthtelling that dominates the marginal incentive to deviate arising from profiles (x_H, y_H, z_H) and (x_H, z_H, y_H). Hence, the TS equilibrium is compatible with conflict profiles (x_H, y_H, z_H) and (x_H, z_H, y_H) and therefore $\hat{q}_D(q_H) > \beta(x_H, y_H, z_H)$.

The role of the consensus effect is key to understanding why unanimous voting rules typically have to be excluded from our analysis. Suppose the voting rule is given by $k = 1$. The consensus effect incentivizes a dove holding some g-signal not to misreport it as an i-signal to avoid potentially undesired acquittals. Now suppose the following sequence of events: first, the dove indeed misreported his g-signal as an i-signal. Secondly, after the communication stage, an acquittal is the outcome triggered by the realized report profile on equilibrium path. Thirdly, given other jurors' truthful reports and his true signal, the dove favors a conviction. Then, the dove can individually impose this desired conviction simply by voting for it. Hence, under unanimity, there is no potential downside of lying for a dove. The same argument appears in Austen-Smith and Feddersen (2006). However, note that their impossibility result with respect to unanimity does not apply in full generality to our environment. Indeed, in the example of Section 3 the TS equilibrium continues to exist if the committee consists of two hawks and one dove and the voting rule is changed to $k = 3$ (unanimity).

Theorem 2 complements Theorem 1 by considering the case of arbitrary numbers of conflict profiles and performing a quantitative analysis of the above mentioned trade-off.

Theorem 2. *Let hawks have critical mass and fix some $m \in \mathbb{N}$ and some $q_H \in (0, 1)$. Then there exists a threshold committee size $\hat{n}$ s.t. for any committee size $n \geq \hat{n}$ and any signal quality $p \geq \frac{1}{2}$, the pair (q_H, q_D) with $q_D = \hat{q}_D(q_H)$ implies at least m conflict profiles.*

Theorem 2 states that the number of conflict profiles compatible with the TS equilibrium is arbitrarily large, if committee size is sufficiently large. The only added requirement with respect to Theorem 1 is a moderate lower bound on signal precision. The existence of the TS equilibrium for a given number of conflict profiles is guaranteed uniformly over signal precisions and large committee sizes. By an additional compactness argument, the statement also applies uniformly over values of q_H located within the $[\epsilon, 1-\epsilon]$-interval, for any $\epsilon > 0$.

We give an intuitive outline of the proof of Theorem 2 in what follows; a formal treatment of technical aspects can be found in the Appendix. In a first step, recall some of the structural results we derived in Section 4: for $p \geq \frac{1}{2}$, signal profiles are ordered in a lexicographic way. The fundamental characteristic of a given signal profile is the relative likelihood of the most likely modality of guilt relative to the likelihood of innocence, measured by $\delta_1(x, y, z) := \max\{y, z\} - x$. The larger δ_1, the larger the posterior probability of guilt. Profiles with identical δ_1-values are ordered according to $\delta_2(x, y, z) := \min\{y, z\} - x$, higher values again indicating higher posteriors for guilt. Limit points of posterior probabilities of guilt are thus spread over the $[0, 1]$-interval and given as $\lim_{x \to \infty} \beta(x, 0, x + \delta_1)$, for any $\delta_1 \in \mathbb{Z}$.

The second step of the proof relates these structural properties of the signal space to the question of pivotality in a putative TS equilibrium. Fix some committee size n and let hawks have critical mass. Suppose a signal profile (x, y, z)

at which a deviation by a dove from a truthful g_2-report to an i-report will change the outcome. Such a signal profile must satisfy
$\beta(x+1, y, z-1) < \beta(x_H, y_H, z_H) \leq \beta(x, y, z)$. All ex post probabilities of guilt arising from signal profiles (x, y, z) that satisfy the above inequalities are necessarily placed within the interval $[\beta(x_H, y_H, z_H), \beta(x_H - 1, y_H, z_H + 1))$. Note that this latter interval typically contains two limit points of ex post probabilities (see Figure 2).

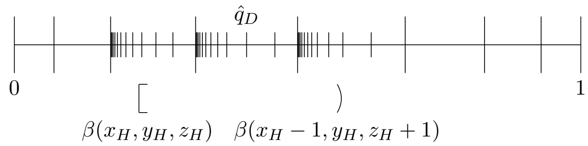

Figure 2: Limit points of ex post probabilities.

The main intuition for Theorem 2 is now the following: the critical dove type $\hat{q}_D(q_H)$ must be located in between q_H and $\beta(x_H - 1, y_H, z_H + 1)$. Indeed, all pivotal profiles generating ex post probabilities of guilt smaller than $\hat{q}_D(q_H)$ incentivize deviating while those generating probabilities larger than this threshold incentivize truthtelling. If the indifference point $\hat{q}_D(q_H)$ is to the left of both limit points, for large n a "double infinite" number of profiles that incentive truthtelling must be exactly counterbalanced by a bounded number of profiles that incentivize deviating, which is implausible. It follows that, for large n, $\hat{q}_D(q_H)$ must lie in between and bounded away from both limit points. But the existence of one limit point of pivotal profiles to the left of $\hat{q}_D(q_H)$ implies that q_H and $\hat{q}_D(q_H)$ determine an arbitrarily large number of conflict profiles when $n \to \infty$. This is the core statement of Theorem 2. As an immediate consequence of the same reasoning, we can conclude that whenever q_H and q_D are within the same interval between limit points, a sufficiently large committee size n is a sufficient condition for the existence of the TS equilibrium. This potentially supportive effect of large committee size for the existence of the TS equilibrium contrasts with the results of Meirowitz (2007), among others.

Theorem 2 provides the general insight that the number of conflict profiles, in our model, has no fundamental bearing upon the existence of the TS equilibrium. In our environment, it therefore does not provide a relevant measure of heterogeneity and conflict among jurors. This insight constrasts with existing results in the literature on deliberation in heterogeneous committees. In Coughlan (2000) and Van Weelden (2008), a single conflict profile already precludes the existence of the TS equilibrium. Le Quement (2012) shows that uncertainty about juror preferences does not suffice to support the TS equilibrium in large committees in the presence of more than one conflict profile.

The analysis of the case where doves have critical mass is qualitatively identical to the above. Equivalents to Theorems 1 and 2 for the case where doves have critical mass are provided in the Appendix as Theorem 1.A and Theorem 2.A. As a closing remark, numerical examples show that the attribution of critical mass matters. Furthermore, no set of voting rules assigning critical mass to the one preference type unilaterally dominates in terms of compatibility with the TS equilibrium. The relevance of the voting rule for the existence of the TS equilibrium is an innovative feature of our model as compared to Coughlan (2000).

6 General committees and first best outcomes

In this section, we release the assumption of only two preference types and sincere voting. Our previous results are shown to carry over to this new environment. This more general setup furthermore allows us to address the question of the implementability of the first best decision rule.

General committees. A natural generalization of the basic hawks vs doves setup is to allow for individual preference parameters $q_j \in (0,1)$ for each juror j. Without loss of generality, assume $q_1 \leq \ldots \leq q_n$, i.e. juror 1 is the harshest juror and juror n the most lenient.

In what follows, we call decision rule $\tilde{q}$ the rule that assigns to each profile of pooled signals the decision that a hypothetical juror of preference type $\tilde{q}$ favors. The equilibrium that we consider in what follows is of the following type: jurors truthfully announce their signals in the communication stage. Subsequently, each juror, independently of his own preference type q_j, votes for conviction if and only if the pooled information is such that a juror of type $\tilde{q}$ would favor a conviction. Accordingly, an individual juror is never pivotal in the voting stage as long as the voting rule is non-unanimous.

The next theorem follows from our characterizations of the previous section.

Theorem 3. *Fix $q_1, ..., q_n$ with $q_1 \leq \ldots \leq q_n$, and some voting rule $k \in \{2, ..., n-1\}$. There exists a truthful equilibrium implementing decision rule $\tilde{q}$ if and only if $q_1 \geq \hat{q}_H(\tilde{q})$ and $q_n \leq \hat{q}_D(\tilde{q})$.*

Theorem 3 states that a truthful equilibrium implementing decision rule $\tilde{q}$ exists if and only if the thresholds of the most extreme jurors q_1 and q_n are within the interval $[\hat{q}_H(\tilde{q}), \hat{q}_D(\tilde{q})]$ defined by the threshold of a virtual juror of preference type $\tilde{q}$. Indeed, if this is the case, it follows trivially that $q_2, ..., q_{n-1}$ are also located within this interval. As compared to the TS equilibrium in the binary setup with hawks and doves, this generalization typically renders truthful communication compatible with a substantially larger spread of preference parameters within the committee. Indeed, in the original setup, the implemented decision rule is that of an extreme preference type (hawks or doves) whereas here, the implemented decision rule is typically that of a hypothetical intermediate preference type.

A main implication of Theorem 3 is as follows: given a set of heterogeneous individuals, all non-unanimous voting rules are equivalent in terms of their compatibility with a truthful equilibrium implementing decision rule $\tilde{q}$. This result is reminiscent of Gerardi and Yariv (2007).

We visualize Theorem 3 in Figure 3 for the classical jury size of $n = 12$ with $p = 0.8$. Given any decision rule $\tilde{q}$ (on the horizontal axis), the dashed (solid) graph indicates the most lenient (harshest) juror type compatible with the existence of a truthful equilibrium implementing decision rule $\tilde{q}$. For every $\tilde{q}$, a truthful equilibrium implementing decision rule $\tilde{q}$ exists if and only if all juror types are located within the interval defined by the two step functions. For our choice of parameters, the number of conflict profiles between the two extreme types defining this interval is typically around 10.

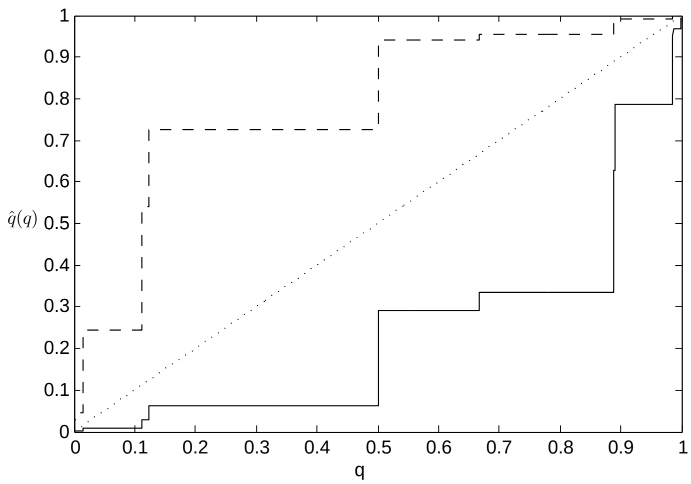

Figure 3: Preference types compatible with truthful equilibrium implementing decision rule q.

Welfare-maximizing outcomes. As demonstrated in Wolinsky (2002), Gerardi and Yariv (2007) and Gerardi, McLean and Postlewaite (2009), conflict in committees typically precludes implementation of the ex post welfare maximizing outcome if utility is non-transferable. This impossibility result frequently breaks down within our environment.

Suppose a committee with individual preference parameters $q_1 \leq \ldots \leq q_n$ and suppose a social planner or a designer wants to maximize the ex post expected utility among jurors according to $Eu = E\sum_{j=1}^{n} \lambda_j u_j = \sum_{j=1}^{n} \lambda_j E u_j$ for given Pareto-weights $\lambda_1, ..., \lambda_n \geq 0$ with $\sum_{j=1}^{n} \lambda_j = 1$. Given fully pooled private signals, the expected ex post utility of juror j conditional on signal profile (x, y, z) is given by

$$Eu_j\left(\left(x,y,z\right),\alpha\right) = \begin{cases} -P\left[\omega \in G|\left(x,y,z\right)\right] \cdot \left(1-q_j\right) & \alpha = A \\ -P\left[\omega = I|\left(x,y,z\right)\right] \cdot q_j & \alpha = C, \end{cases}$$

so

$$Eu((x,y,z),\alpha) = \begin{cases} -P[\omega \in G|(x,y,z)] \cdot \left(1-\sum_{j=1}^{n}\lambda_j q_j\right) & \alpha = A \\ -P[\omega = I|(x,y,z)] \cdot \sum_{j=1}^{n}\lambda_j q_j & \alpha = C. \end{cases}$$

Hence the socially optimal action, given fully pooled information, coincides with the preferred action of a hypothetical juror with preference parameter $\overline{q} = \sum_{j=1}^{n}\lambda_j q_j$.

Theorem 4. *Fix preference parameters $q_1, ..., q_n$, with $q_1 \leq ... \leq q_n$, and define $\overline{q} = \sum_{j=1}^{n}\lambda_j q_j$. The welfare maximizing decision rule with respect to $Eu = \sum_{j=1}^{n}\lambda_j Eu_j$ is implementable if and only if $\overline{q}$ satisfies the assumptions of Theorem 3.*

Theorem 4 provides a simple criterion for the implementability of the welfare maximizing decision rule in general committees. Consider a mediator who receives private reports from all the jurors and recommends to all jurors to vote for conviction if and only if a juror of preference type $\overline{q}$ would favor a conviction. Suppose that in this mediation game truthful reporting and obedient voting are incentive compatible. The incentives of a juror in a truthful equilibrium implementing the decision rule $\overline{q}$ are identical to those that he faces in an equilibrium of the mediation game that implements decision rule $\overline{q}$. By the revelation principle, if the decision rule $\overline{q}$ can be implemented, it can be implemented in the mediation game and thus in a truthful equilibrium. This, together with Theorem 3, proves Theorem 4. An attractive feature of Theorem 4 is that the welfare maximizing decision rule can be implemented without resorting to an outside agent endowed with commitment power.

On the other hand, the alternative approach of installing a mediator who decides according to some arbitrary preference type $\tilde{q}$ and recommends unanimous voting is compatible with the assumption of sincere voting. Indeed, as long as the mediator does not publicly reveal the collected signals, all jurors will be willing to follow his recommendation at the voting stage if they were willing to report their information truthfully at the communication stage. Hence, the conditions in Theorem 3 ensure the incentive compatibility of truthful revelation followed by sincere voting in the mediation game.

7 Bridging the gap to the classical Condorcet model

In this final section, we show that the simultaneous addition of two ingredients is key to determining the breakdown of Coughlan's impossibility result: a richer message space as well as a richer signal structure. We first briefly examine the role of an increased message space and subsequently analyze in more detail the role of a richer signal structure.

Suppose that the message space is of cardinality two instead of three. Denote this space by $\{i, g\}$. An example of such an environment would be straw polls in which jurors first cast a non-binding vote and subsequently, upon observation of the aggregate profile of straw votes, cast a binding vote. In such an environment, one may ask whether there exists an equilibrium of the following type. First, in the communication (straw poll) stage, jurors send message i when holding an i-signal while they send message g when holding either type of g-signal. In the second stage, jurors vote sincerely.

In the above putative equilibrium, it is easy to see that the total number of g-reports revealed is all jurors take into account when casting a sincere vote. Suppose that the committee is heterogeneous in the sense that for at least one number of g-reports, there is disagreement between hawks and doves about the optimal decision. Now suppose hawks have critical mass. Then, the report of a dove is pivotal if and only if the number of guilty signals reported by other jurors is exactly one unit less than the number of guilty signals that hawks require to prefer conviction. But in this case, a dove holding some g-signal still prefers acquittal in expectation and hence will report an i-signal to implement his preferred outcome. This is precisely the mechanism that drives Coughlan's impossibility result.

Next, we analyze the role played by a richer signal structure. We proceed by explicitly linking our model to the standard Condorcet model in which signals only contain information about whether the defendant is innocent or guilty. Thereby, we show that the Condorcet model is nested within a generalized version of our model and we once more highlight the role played by consistency as the fundamental difference between both setups.

Recall the information structure of our analysis so far: signals are i.i.d. conditional on the realized state of the world $\omega \in \{I, G_1, G_2\}$; they show the correct state of the world with probability $p \in \left(\frac{1}{3}, 1\right)$ while indicating either of the remaining states with probability $p_r = \frac{(1-p)}{2}$, the subscript r standing for "residual". Call this information structure ungarbled.

Suppose now instead a garbled information structure that fully eliminates the informational distinction between the two modalities of guilt: signals do no longer contain any information about whether the true state of the world is G_1 or G_2. Let the signal generating process take the following form: if the true state is I, an i-signal is generated with probability p, while either type of guilty signal (g_1 or g_2) is generated with probability $\frac{p_r+p_r}{2} = p_r$. If either of the guilty states is the true state of the world, an i-signal is generated with probability p_r while a g_1- or g_2-signal is generated with probability $\frac{p+p_r}{2}$.

Next, we can consider arbitrary mixtures of these two information structures via an additional parameter $\lambda \in [0, 1]$. A λ-mixed signal generating process attaches weight λ to the ungarbled signal generating process and weight $1 - \lambda$ to the garbled signal generating process. Formally, if I is the true state of the world, an i-signal is generated with probability p. On the other hand, each of the guilty-signals is generated with probability $\lambda \cdot p_r + (1 - \lambda) \cdot \frac{p_r+p_r}{2} = p_r$. Note that none of these numbers depends on λ. However, if e.g. G_1 is the true state of the world, an i-signal is generated with probability p_r, a g_1-signal is generated with probability $p_\lambda := \lambda \cdot p + (1 - \lambda) \cdot \frac{p+p_r}{2}$ and a g_2-signal is generated with probability $p_{r,\lambda} := \lambda \cdot p_r + (1-\lambda) \cdot \frac{p+p_r}{2}$. Here, the probabilities of different signals occurring clearly depend on λ. The case of G_2 being the true state of the world is symmetric.

A λ-mixed information structure is not merely a technical construct but potentially captures some realistic features of information which we have so far abstracted from. Some bits of a juror's signal may well refer to guilt without explicitly favoring a particular modality. As an example, the presence of the defendant's fingerprints at the scene of the crime typically does not convey information about the points in time at which the crime could have been committed, which constitute different modalities. However, as long as $\lambda > 0$, we stick to the assumption that some bits of information contained in a given signal indeed refer to a particular modality of guilt.

Given a $\lambda-$mixed signal generating process and a signal precision p, the Bayesian posterior probability of guilt given signal profile (x, y, z) is given as

$$\beta_\lambda (x, y, z) \quad = \quad \frac{p_r^x \cdot p_\lambda^y \cdot p_{r,\lambda}^z + p_r^x \cdot p_{r,\lambda}^y \cdot p_\lambda^z}{p^x \cdot p_r^y \cdot p_r^z + p_r^x \cdot p_\lambda^y \cdot p_{r,\lambda}^z + p_r^x \cdot p_{r,\lambda}^y \cdot p_\lambda^z}.$$

Using the coordinates introduced in Section 4, namely $n = x+y+z$, $n_g = y+z$ and $\Delta = |z-y|$, the following lemma provides a direct generalization of Lemma 1:

Lemma 5. *For fixed n and n_g, $\beta_\lambda = \beta_\lambda (n, n_g, \Delta)$ is weakly increasing in Δ for any $\lambda \geq 0$, and strictly increasing in Δ if and only if $\lambda > 0$.*

Lemma 5 shows that, unless signals do not transmit *any* information about the modality of guilt ($\lambda = 0$), increasing consistency among a given total number of guilty signals increases the posterior probability of guilt. This implies that, although potentially mitigated, the driving forces of our results, namely the consensus effect and the uncertainty effect, remain present for any positive λ. In other words, the qualitative insights of our main analysis remain valid except for the limit case where no informational distinction at all can be made between different modalities of guilt.

If $\lambda = 0$, it follows that $p_0 = p_{r,0} = \frac{p+p_r}{2}$ and after some algebra β_0 simplifies to

$$\beta_0 (n, n_g, \Delta) \quad = \quad \frac{\frac{2}{3} \cdot \left(\frac{1-p}{2}\right)^{n-n_g} \cdot \left(\frac{1+p}{2}\right)^{n_g}}{\frac{1}{3} \cdot p^{n-n_g} \cdot (1-p)^{n_g} + \frac{2}{3} \cdot \left(\frac{1-p}{2}\right)^{n-n_g} \cdot \left(\frac{1+p}{2}\right)^{n_g}}.$$

In particular, β_0 is increasing in n_g and does not depend on Δ, in line with Lemma 5. Our model is now mathematically equivalent to the classical two states and two signals Condorcet model with the following specifications: the defendant is guilty with prior probability $\frac{2}{3}$. If the defendant is guilty, a g-signal is generated with probability $\frac{1+p}{2}$ and an i-signal is generated with probability $\frac{1-p}{2}$. In case the defendant is innocent, an i-signal is generated with probability p and a g-signal is generated with probability $1-p$. It follows that Coughlan's impossibility result applies.

8 Conclusion

We find that in a simple jury model with pre-vote communication that accounts for the role of informational consistency, substantial preference divergence among jurors is frequently compatible with the existence of the TS equilibrium. Furthermore, we identify the driving forces underlying this result, namely the consensus and uncertainty effects, both of which originate in the emerging multiplicity of pivotal scenarios faced by jurors in the communication stage. We subsequently present conditions for the implementability of first best decision rules through truthful equilibria of our game. These conditions are satisfied for a large set of parameter values and independent of the chosen (non-unanimous) voting rule.

The consistency concern provides an innovative approach to the general issue of the contextual determination of meaning in communication games. That is, the relative impact of given statements is often to a large extent dependent on the remaining information available to the audience. In our environment, announcements exhibit positive complementarity. We believe that the contextual generation of meaning, whether featuring positive or negative complementarity, could fruitfully be studied within the context of other communication games.

A central aspect of our analysis is the asymmetry characterizing our environment: simple decisions combine with rich information and rich language. Our results indicate that rich information and language can facilitate information pooling between heterogeneous agents. These features generate communicational complexity that fosters the appearance of trade-offs in truthtelling incentives. We consider this insight worth investigating more.

Appendix

Proof of Lemma 1. Lemma 1 is a special case of Lemma 5 whose proof is found at the end of the Appendix. □

Proof of Lemma 2. To simplify notation, define $\gamma := \frac{2p}{1-p}$. Then $\beta(\delta_1, \delta_2) = \frac{\gamma^{\delta_1}+\gamma^{\delta_2}}{1+\gamma^{\delta_1}+\gamma^{\delta_2}}$ is increasing in δ_1 and δ_2. Now compare $\beta(\delta_1, \delta_2)$ and $\beta\left(\tilde{\delta}_1, \tilde{\delta}_2\right)$ with $\delta_1 > \tilde{\delta}_1$. As by definition $\delta_1 \geq \delta_2$, $\tilde{\delta}_1 \geq \tilde{\delta}_2$, and as $p \geq \frac{1}{2}$ implies $\gamma \geq 2$, we have

$$\begin{aligned} \beta(\delta_1, \delta_2) &= \frac{\gamma^{\delta_1}+\gamma^{\delta_2}}{1+\gamma^{\delta_1}+\gamma^{\delta_2}} > \frac{\gamma^{\delta_1}}{1+\gamma^{\delta_1}} \geq \frac{\gamma^{\tilde{\delta}_1+1}}{1+\gamma^{\tilde{\delta}_1+1}} \\ &\geq \frac{\gamma^{\tilde{\delta}_1}+\gamma^{\tilde{\delta}_1}}{1+\gamma^{\tilde{\delta}_1}+\gamma^{\tilde{\delta}_1}} \geq \frac{\gamma^{\tilde{\delta}_1}+\gamma^{\tilde{\delta}_2}}{1+\gamma^{\tilde{\delta}_1}+\gamma^{\tilde{\delta}_2}} = \beta\left(\tilde{\delta}_1, \tilde{\delta}_2\right). \end{aligned}$$

This proves the lexicographic structure claimed by the Lemma. □

Proof of Lemma 3. Any pair $(\delta_1, \delta_2) \in \mathbb{Z}^2$ with $\delta_1 \geq \delta_2$ can arise from a signal profile $(x, y, z) \in \mathbb{N}^3$ by fixing some natural number $l \geq |\delta_2|$ and setting $x = l$, $y = l + \delta_2$, $z = l + \delta_1$.

To see that all limit points are of the claimed form, take an arbitrary convergent sequence $\beta(x_n, y_n, z_n)$, $n \in \mathbb{N}$, that does not have a constant subsequence. If $\delta_1(x_n, y_n, z_n) \to \pm\infty$, then $\beta(x_n, y_n, z_n)$ converges to 1 or 0. Otherwise, by the lexicographic ordering, $\delta_1(x_n, y_n, z_n)$ ultimately becomes stationary, and by convergence, $\delta_2(x_n, y_n, z_n) \to -\infty$ as $\delta_2(x_n, y_n, z_n)$ is bounded from above by $\delta_1(x_n, y_n, z_n)$. This proves the Lemma. $\square$

Proof of Lemma 4. Suppose a juror with preference parameter q holding a g_2-signal and let (x, y, z) be an arbitrary profile of the remaining $n-1$ jurors with $y \leq z$. If both g-reports lead to the same outcome the profile (x, y, z) provides no incentives with respect to a deviation across g-signals. In particular, this is the case if $y = z$. If the two g-reports lead to different outcomes, by the assumption of $y \leq z$ a g_2-report will lead to conviction for profile (x, y, z) and to acquittal for the transposed profile (x, z, y), while a g_1-report will lead to conviction for profile (x, z, y) and to acquittal for profile (x, y, z). Comparing expected utilities from a g_1- and a g_2-report conditional on the remaining jurors' signal profile being either (x, y, z) or (x, z, y) yields

$$
\begin{aligned}
& Eu\left[g_2 |\left\{(x,y,z),(x,z,y)\right\}\right] - Eu\left[g_1 |\left\{(x,y,z),(x,z,y)\right\}\right] \\
= \; & P\left[I|g_2\right] \cdot \left(P\left[(x,z,y)|I\right] - P\left[(x,y,z)|I\right]\right) \cdot q \\
& + P\left[G_1|g_2\right] \cdot \left(P\left[(x,y,z)|G_1\right] - P\left[(x,z,y)|G_1\right]\right) \cdot (1-q) \\
& + P\left[G_2|g_2\right] \cdot \left(P\left[(x,y,z)|G_2\right] - P\left[(x,z,y)|G_2\right]\right) \cdot (1-q) \\
= \; & \frac{n! \cdot \left(\frac{1-p}{2}\right)^n}{x! \cdot y! \cdot z!} \left(\left(\frac{2p}{1-p}\right) - 1\right) \left(\left(\frac{2p}{1-p}\right)^z - \left(\frac{2p}{1-p}\right)^y\right)(1-q) \\
> \; & 0.
\end{aligned}
$$

Hence, for any pair of transposed profiles, the juror is either indifferent between lying and reporting truthfully or has a strict incentive to report truthfully. Given that the set of feasible signal profiles of other jurors can entirely be fragmented into such pairs, this proves the result. Clearly, the result applies in a symmetric fashion to a juror holding a g_1-signal and considering a deviation towards reporting g_2. $\square$

Lemma A. Fix n, assume $p \geq \frac{1}{2}$ and let hawks have critical mass[1] with threshold profile (x_H, y_H, z_H) such that $\Delta_H = z_H - y_H \geq 0$. Then the set PIV of all pivotal profiles consists of the following profiles of $n-1$ signals:[2]

[1]The case of doves having critical mass is identical, change subscripts from H to D.
[2]For any real number w, $\lfloor w \rfloor$ denotes the largest integer smaller or equal than w.

$$
\begin{aligned}
Piv_1(r) &= (x_H + r, y_H - 2(r+1), z_H + r + 1) \quad \text{for } r = 0, ..., \left\lfloor \frac{y_H}{2} \right\rfloor - 1 \\
Piv_1^T(r) &= (x_H + r, z_H + r + 1, y_H - 2(r+1)) \quad \text{for } r = 0, ..., \left\lfloor \frac{y_H}{2} \right\rfloor - 1 \\
Piv_2(s) &= (x_H - s, y_H + 2s - 1, z_H - s) \quad \text{for } s = 0, ..., \min\left\{ \left\lfloor \frac{\Delta_H + 1}{3} \right\rfloor, x_H \right\} \\
Piv_2^T(s) &= (x_H - s, z_H - s, y_H + 2s - 1) \quad \text{for } s = 0, ..., \min\left\{ \left\lfloor \frac{\Delta_H + 1}{3} \right\rfloor, x_H \right\} \\
Piv_3(\tilde{r}) &= (x_H + \tilde{r}, y_H - 2\tilde{r} - 1, z_H + \tilde{r}) \quad \text{for } \tilde{r} = 0, ..., \left\lfloor \frac{y_H - 1}{2} \right\rfloor \\
Piv_3^T(\tilde{r}) &= (x_H + \tilde{r}, z_H + \tilde{r}, y_H - 2\tilde{r} - 1) \quad \text{for } \tilde{r} = 0, ..., \left\lfloor \frac{y_H - 1}{2} \right\rfloor \\
Piv_4(\tilde{s}) &= (x_H - \tilde{s}, y_H + 2\tilde{s}, z_H - \tilde{s} - 1) \quad \text{for } \tilde{s} = 0, ..., \min\left\{ \left\lfloor \frac{\Delta_H - 1}{3} \right\rfloor, x_H \right\} \\
Piv_4^T(\tilde{s}) &= (x_H - \tilde{s}, z_H - \tilde{s} - 1, y_H + 2\tilde{s}) \quad \text{for } \tilde{s} = 0, ..., \min\left\{ \left\lfloor \frac{\Delta_H - 1}{3} \right\rfloor, x_H \right\}
\end{aligned}
$$

where

$$
\begin{aligned}
Piv_2(0) &= Piv_3(0) \\
Piv_2^T(0) &= Piv_3^T(0) \\
Piv_2\left(\left\lfloor \frac{\Delta_H + 1}{3} \right\rfloor \right) &= Piv_2^T\left(\left\lfloor \frac{\Delta_H + 1}{3} \right\rfloor \right) \quad \text{if } \Delta_H + 1 = 0 \mod 3 \\
Piv_4\left(\left\lfloor \frac{\Delta_H - 1}{3} \right\rfloor \right) &= Piv_4^T\left(\left\lfloor \frac{\Delta_H - 1}{3} \right\rfloor \right) \quad \text{if } \Delta_H - 1 = 0 \mod 3.
\end{aligned}
$$

Proof of Lemma A. Let (x, y, z) be an arbitrary signal profile of $n-1$ jurors. Assume without loss of generality that $z \geq y$. The profile (x, y, z) then is a pivotal profile if and only if

$$\beta(x+1, y, z) < \beta(x_H, y_H, z_H) \leq \beta(x, y, z+1).$$

By Lemma 2 this implies

$$
\begin{cases}
(z - x) = z_H - x_H + 1 \wedge y - x \leq y_H - x_H & \text{or} \\
(z - x) = z_H - x_H & \text{or} \\
(z - x) = z_H - x_H - 1 \wedge y - x \geq y_H - x_H.
\end{cases}
$$

With $n = x_H + y_H + z_H$ fixed, the first case corresponds to profiles of type Piv_1, the second case to profiles of type Piv_2 and Piv_3 and the third case to profiles of type Piv_4. The transposed profiles are derived in exactly the same way, assuming $y \geq z$ and comparing a g_1-report with an i-report. □

Proof of Theorem 1. Without loss of generality, consider a dove holding a g_2-signal. Recall that an i-report leads to acquittal for any pivotal profile and write PIV_{g_2} for the set of all pivotal profiles for which a g_2-report leads to conviction. Note that PIV_{g_2} depends on q_H resp. (x_H, y_H, z_H) only.

a) The TS equilibrium exists if and only if expected utility from reporting g_2 is at least as large as expected utility from reporting i, that is

$$\begin{aligned} Eu(g_2) - Eu(i) &= -\sum_{(x,y,z)\in PIV_{g_2}} P[I|g_2]\cdot P[(x,y,z)|I]\cdot q_D \\ &\quad + \sum_{(x,y,z)\in PIV_{g_2}} P[G|g_2]\cdot P[(x,y,z)|G]\cdot(1-q_D) \\ &\geq 0. \end{aligned}$$

The above expression is decreasing in q_D, hence the inequality holds for any

$$q_D \leq \frac{\sum_{(x,y,z)\in PIV_{g_2}} P[G|g_2]\cdot P[(x,y,z)|G]}{\sum_{(x,y,z)\in PIV_{g_2}} \left(P[G|g_2]\cdot P[(x,y,z)|G] + P[I|g_2]\cdot P[(x,y,z)|I]\right)}$$

and equality yields $\hat{q}_D(q_H)$.

b) $Eu(g_2) - Eu(i)$ is continuous and linearly decreasing in q_D as seen above. Consider the boundary case of essentially homogeneous preferences: for $q_D = \beta(x_H, y_H, z_H)$, doves holding a g_2-signal are indifferent between both outcomes for profile $Piv_4(0) = (x_H, y_H, z_H - 1)$ while for all other pivotal profiles they prefer truthtelling over deviating towards an i-report. To prove strict incentives for truthtelling, it therefore suffices to guarantee the existence of another pivotal profile. We distinguish three cases.

First, if $(x_H, y_H, z_H) = (n, 0, 0)$, hawks will implement conviction independently of all reports, there are no pivotal profiles at all and hence the TS equilibrium trivially exists.

Secondly, if $y_H > 0$, the remaining jurors hold signal profile $Piv_3(0) = (x_H, y_H - 1, z_H)$ with positive probability. For any $p > \frac{1}{3}$, we have $\beta(x_H + 1, y_H - 1, z_H) < \beta(x_H, y_H, z_H) < \beta(x_H, y_H - 1, z_H + 1)$, hence doves have a strict incentive for truthtelling.

Thirdly, if $y_H = 0$ and $x_H > 0$, $z_H \geq 2$, the remaining jurors hold signal profile $Piv_2(1) = (x_H - 1, y_H + 1, z_H - 1)$ with positive probability. Here, for any $p > \frac{1}{3}$ we have $\beta(x_H, y_H + 1, z_H - 1) < \beta(x_H, y_H, z_H) < \beta(x_H - 1, y_H + 1, z_H)$, so again doves have a strict incentive for truthtelling.

As by assumption $(x_H, y_H, z_H) \notin \{(0,0,n), (n-1,0,1)\}$, it follows that doves with preference parameter $q_D = \beta(x_H, y_H, z_H)$ have a strict incentive for truthtelling. The result then follows from continuity and monotonicity of $Eu(g_2) - Eu(i)$ with respect to q_D. □

Proof of Theorem 2. The proof of Theorem 2 relies on the structural insights from Lemmas 2 and 3. Consider without loss of generality a dove holding a g_2-signal. Fix some $q_H \in (0,1)$. Then, one of the following two cases applies:

1. For some jury size $\tilde{n}$, the threshold profile (x_H, y_H, z_H) satisfies $y_H \geq 2$.
2. For any jury size n, the threshold profile (x_H, y_H, z_H) satisfies $y_H < 2$.

Fix m and q_H resp. (x_H, y_H, z_H) for some large committee size n. To abbreviate notation, write $\gamma := \frac{2p}{1-p} \geq 2$. For either of the above cases, we show that there exists a value for q_D that implies at least m conflict profiles and at the same time is compatible with the existence of the TS equilibrium.

The case $y_H \geq 2$:
Note that q_H is necessarily bounded away from any limit point of the set of posterior probabilities as for jury size $\tilde{n}$ we have

$$\frac{\left(\frac{2p}{1-p}\right)^{z_H - x_H}}{1+\left(\frac{2p}{1-p}\right)^{z_H-x_H}} < \beta\left(x_H+1, y_H-2, z_H+1\right) < q_H \leq \beta\left(x_H, y_H, z_H\right) < \frac{\left(\frac{2p}{1-p}\right)^{z_H-x_H+1}}{1+\left(\frac{2p}{1-p}\right)^{z_H-x_H+1}}.$$

It is therefore enough to prove the claim for the following threshold profiles: (x_H, y_H, z_H) with $y_H \geq 2$ and constant $\delta_1 = z_H - x_H$, $\delta_2 = y_H - x_H$, where x_H, y_H, z_H can be arbitrarily large. To ensure the existence of at least m conflict profiles, it suffices to check that the TS equilibrium exists for

$$q_D = \beta\left(x_H + \frac{y_H}{2} - m - 1, 2m+1, z_H + \frac{y_H}{2} - m\right) = \frac{\gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}}{\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}}$$

with $y_H \geq 2m+2$. This choice guarantees the existence of at least m conflict profiles, namely profiles $Piv_3\left(r\right) + (0,0,1)$ for $r = \left\lfloor \frac{y_H - 1}{2} \right\rfloor - m - 1, ..., \left\lfloor \frac{y_H-1}{2} \right\rfloor$ (see Lemma A for the definition of $Piv_3\left(r\right)$).

We now show that the overall sum of incentives for truthtelling arising from these profiles is positive. We do so by combining detrimental profiles with suitable beneficial profiles and show that the previous statement holds for each such combination individually.

Combining profiles $Piv_2^T(s)$ and $Piv_4(s)$ for $s \geq 0$ yields

$$\begin{aligned}
& Eu(g_2|Piv_2^T(s)) - Eu(i|Piv_2^T(s)) + Eu(g_2|Piv_4(s)) - Eu(i|Piv_4(s)) \\
= \; & -P\left[I|g_2\right] \cdot P\left[\left(x_H - s, z_H - s, y_H + 2s - 1\right)|I\right] \cdot q_D \\
& + P\left[G_1|g_2\right] \cdot P\left[\left(x_H - s, z_H - s, y_H + 2s - 1\right)|G_1\right] \cdot (1 - q_D) \\
& + P\left[G_2|g_2\right] \cdot P\left[\left(x_H - s, z_H - s, y_H + 2s - 1\right)|G_2\right] \cdot (1 - q_D) \\
& - P\left[I|g_2\right] \cdot P\left[\left(x_H - s, y_H + 2s, z_H - s - 1\right)|I\right] \cdot q_D \\
& + P\left[G_1|g_2\right] \cdot P\left[\left(x_H - s, y_H + 2s, z_H - s - 1\right)|G_1\right] \cdot (1 - q_D) \\
& + P\left[G_2|g_2\right] \cdot P\left[\left(x_H - s, y_H + 2s, z_H - s - 1\right)|G_2\right] \cdot (1 - q_D) \\
= \; & \frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{(x_H - s)! \cdot (y_H + 2s)! \cdot (z_H - s)! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)} \\
& \cdot \left(z_H + y_H + s\right) \cdot \left(-\gamma^{-s+2m+1} - \gamma^{z_H + \frac{y_H}{2} - s - m} + \gamma^{\frac{3}{2}y_H + 2s - m - 1} + \gamma^{z_H + \frac{y_H}{2} - s - m - 1}\right).
\end{aligned}$$

For fixed q_H and n sufficiently large, x_H, y_H, z_H are large relative to the differences between each other as well as relative to s. The first line then barely depends on s while the second line is increasing in s. So, for any feasible s,

$$\begin{aligned}
& Eu(g_2|Piv_2^T(s)) - Eu(i|Piv_2^T(s)) + Eu(g_2|Piv_4(s)) - Eu(i|Piv_4(s)) \\
\geq \; & Eu(g_2|Piv_2^T(0)) - Eu(i|Piv_2^T(0)) + Eu(g_2|Piv_4(0)) - Eu(i|Piv_4(0)).
\end{aligned}$$

To outweigh the potential loss from profiles $Piv_2^T(s)$ and $Piv_4(s)$, consider profiles $Piv_1(r)$ for $r = 0, ..., 2\left\lfloor \frac{\Delta_H + 1}{3} \right\rfloor + 1$, where $2\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor + 1 \leq \left\lfloor \frac{y_H - 1}{2} \right\rfloor - m - 3$

for sufficiently large n so that y_H is sufficiently large. Note that

$$\begin{aligned} & Eu(g_2|Piv_1(r)) - Eu(i|Piv_1(r)) \\ = \quad & \frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{(x_H+r)! \cdot (y_H - 2\,(r+1))! \cdot (z_H + r + 1)! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)} \\ & \cdot \left(-\gamma^{r+2m+1} - \gamma^{z_H + \frac{y_H}{2} + r - m} + \gamma^{\frac{3}{2} y_H - 2(r+1) - m - 1} + \gamma^{z_H + \frac{y_H}{2} + r - m + 1}\right). \end{aligned}$$

As before, for n sufficiently large, x_H, y_H, z_H are large relative to r and hence the second line barely depends on r while the third line is increasing in r. So, for any r under consideration,

$$Eu(g_2|Piv_1(r)) - Eu(i|Piv_1(r)) \geq Eu(g_2|Piv_1(0)) - Eu(i|Piv_1(0)).$$

Summing up yields

$$\begin{aligned} & \sum_{s=0}^{\lfloor \frac{\Delta_H+1}{3} \rfloor} Eu(g_2|Piv_2^T(s)) - Eu(i|Piv_2^T(s)) + Eu(g_2|Piv_4(s)) - Eu(i|Piv_4(s)) \\ & + \sum_{r=0}^{2\lfloor \frac{\Delta_H+1}{3} \rfloor} Eu(g_2|Piv_1(r)) - Eu(i|Piv_1(r)) \\ \geq \quad & \left(\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor + 1\right) \left(Eu(g_2|Piv_2^T(0)) - Eu(i|Piv_2^T(0)) + Eu(g_2|Piv_4(0)) - Eu(i|Piv_4(0))\right) \\ & + 2\left(\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor + 1\right) \left(Eu(g_2|Piv_1(0)) - Eu(i|Piv_1(0))\right) \\ = \quad & \frac{\left(\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor + 1\right) \cdot (z_H + y_H) \cdot (n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{x_H! \cdot y_H! \cdot z_H! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)} \\ & \cdot \left(-\gamma^{2m+1} - \gamma^{z_H + \frac{y_H}{2} - m} + \gamma^{\frac{3}{2} y_H - m - 1} + \gamma^{z_H + \frac{y_H}{2} - m - 1}\right) \\ & + \frac{\left(\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor + 1\right) \cdot 2 \cdot (n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{x_H! \cdot (y_H - 2)! \cdot (z_H + 1)! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)} \\ & \cdot \left(-\gamma^{2m+1} - \gamma^{z_H + \frac{y_H}{2} - m} + \gamma^{\frac{3}{2} y_H - 2 - m - 1} + \gamma^{z_H + \frac{y_H}{2} - m + 1}\right). \end{aligned}$$

The factor in front of the first polynomial in γ is approximately as large as the factor in front of the second polynomial, given that n and hence x_H, y_H, z_H are sufficiently large. As for large n we have $y_H \gg m$, concerning the polynomials we conclude that

$$\begin{aligned} & -\gamma^{2m+1} - \gamma^{z_H + \frac{y_H}{2} - m} + \gamma^{\frac{3}{2} y_H - m - 1} + \gamma^{z_H + \frac{y_H}{2} - m - 1} \\ & -\gamma^{2m+1} - \gamma^{z_H + \frac{y_H}{2} - m} + \gamma^{\frac{3}{2} y_H - 2 - m - 1} + \gamma^{z_H + \frac{y_H}{2} - m + 1} \\ \geq \quad & (\gamma - 1) \cdot \left(+\gamma^{z_H + \frac{y_H}{2} - m} - \gamma^{z_H + \frac{y_H}{2} - m - 1}\right) \\ > \quad & 0. \end{aligned}$$

Summing up incentives from profiles $Piv_1(r-1)$, $Piv_1^T(r-1)$ and $Piv_3(r)$ for $r = \lfloor \frac{y_H - 1}{2} \rfloor - m - 2, ..., \lfloor \frac{y_H - 1}{2} \rfloor$ with $y_H \gg m$ yields

$$
\begin{aligned}
& \frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{(x_H+r)! \cdot (y_H-2r)! \cdot (z_H+r)! \cdot \left(\gamma^{x_H+\frac{y_H}{2}-m-1}+\gamma^{2m+1}+\gamma^{z_H+\frac{y_H}{2}-m}\right)} \\
& \cdot \Big(+(x_H+r)\,\gamma^{z_H+\frac{y_H}{2}+r-m}-(x_H+r)\,\gamma^{z_H+\frac{y_H}{2}+r-m-1}-(y_H-2r)\,\gamma^{r+2m+1} \\
& -2\,(x_H+r)\,\gamma^{r+2m}+(x_H+r)\,\gamma^{\frac{3}{2}y_H-2r-m}+(x_H+r)\,\gamma^{\frac{3}{2}y_H-2r-m-1}+(y_H-2r)\,\gamma^{\frac{3}{2}y_H-2r-m-2}\Big) \\
> \quad & \frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H} \cdot \gamma^{r}}{(x_H+r)! \cdot (y_H-2r)! \cdot (z_H+r)! \cdot \left(\gamma^{x_H+\frac{y_H}{2}-m-1}+\gamma^{2m+1}+\gamma^{z_H+\frac{y_H}{2}-m}\right)} \\
& \cdot \left(+(x_H+r)\,\gamma^{z_H+\frac{y_H}{2}-m}-(x_H+r)\,\gamma^{z_H+\frac{y_H}{2}-m-1}-(y_H-2r)\,\gamma^{2m+1}-2\,(x_H+r)\,\gamma^{2m}\right) \\
> \quad & 0.
\end{aligned}
$$

For y_H even, summing up incentives from profiles $Piv_1(r)$, $Piv_1^T(r)$ with $r = \frac{y_H}{2} - 1$ and $y_H \gg m$ yields

$$
\begin{aligned}
& \frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{(x_H+r)! \cdot (y_H-2r-2)! \cdot (z_H+r+1)! \cdot \left(\gamma^{x_H+\frac{y_H}{2}-m-1}+\gamma^{2m+1}+\gamma^{z_H+\frac{y_H}{2}-m}\right)} \\
& \cdot \left(+\gamma^{z_H+y_H-m}-\gamma^{z_H+y_H-m-1}-2\gamma^{\frac{y_H}{2}+2m}+\gamma^{\frac{y_H}{2}-m}+\gamma^{\frac{y_H}{2}-m-1}\right) \\
> \quad & 0.
\end{aligned}
$$

Finally, by choice of q_D all remaining profiles incentivize truthtelling:

$Piv_1(r)$ for all $r = 2\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor + 2, ..., \left\lfloor \frac{y_H-1}{2} \right\rfloor - m - 4$,

$Piv_1^T(r)$ for all $r = 0, ..., \left\lfloor \frac{y_H-1}{2} \right\rfloor - m - 4$,

$Piv_2(s)$ for all $s \geq 0$ given that $y_H \geq 2m+2$,

$Piv_3(r)$ for all $r = 0, ..., \left\lfloor \frac{y_H-1}{2} \right\rfloor - m - 3$.

The case $y_H < 2$:

In this case, we have $q_H \in \left(\beta(x_H+1, z_H, z_H), \frac{\left(\frac{2p}{1-p}\right)^{z_H-x_H}}{1+\left(\frac{2p}{1-p}\right)^{z_H-x_H}}\right]$. Consider therefore threshold profiles (x_H, y_H, z_H) with $y_H < 2$ and constant $\delta_1 = z_H - x_H$, where z_H and x_H can be arbitrarily large.

To ensure the existence of at least m conflict profiles, it suffices to check that the TS equilibrium exists for

$$
q_D = \beta\left(x_H - \left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor, y_H + 2\left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor - 4, z_H - \left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor\right) = \frac{\gamma^{y_H + 3\left\lfloor \frac{z_H-y_H-1}{3} \right\rfloor - 4} + \gamma^{z_H}}{\gamma^{x_H} + \gamma^{y_H + 3\left\lfloor \frac{z_H-y_H-1}{3} \right\rfloor - 4} + \gamma^{z_H}}
$$

with y_H being either 0 or 1. This choice guarantees conflict for any profile of type $Piv_4(s) + (0,0,1)$ with $s \leq \left\lfloor \frac{z_H-y_H-1}{3} \right\rfloor - 2$, and since n and therefore z_H is arbitrarily large while $y_H < 2$, this provides at least m conflict profiles. Note that no pivotal profiles of type Piv_1, Piv_1^T or Piv_3, Piv_3^T exist (except possibly $Piv_3(0) = Piv_2(0)$, $Piv_3^T(0) = Piv_2^T(0)$).

By similar arguments as developed in the case of $y_H \geq 2$, profiles $Piv_2(0)$, $Piv_2^T(0)$ (if they exist) as well as profiles $Piv_2(s), Piv_2^T(s), Piv_4(s-1)$ for $1 \leq s \leq$

$\left\lfloor \frac{\Delta_H - 1}{3} \right\rfloor$ sum up to a positive value for sufficiently large n. Furthermore, profiles $Piv_2\left(\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor\right)$, $Piv_2^T\left(\left\lfloor \frac{\Delta_H+1}{3} \right\rfloor\right)$, $Piv_4\left(\left\lfloor \frac{\Delta_H-1}{3} \right\rfloor\right)$ incentivize truthtelling. $\square$

Theorem 1.A. *Let doves have critical mass.*

a) For any dove type q_D the TS equilibrium exists if and only if the value of q_H lies above a given lower bound $\hat{q}_H(q_D) < q_D$.

b) For any q_D s.t. $(x_D, y_D, z_D) \notin \{(n,0,0), (n-1,0,1)\}$, the pair (q_H, q_D) with $q_H = \hat{q}_H(q_D)$ implies at least one conflict profile.

Proof of Theorem 1.A. The proof of Theorem 1.A is virtually identical to the proof of Theorem 1 and therefore omitted. $\square$

Theorem 2.A. *Let doves have critical mass and fix some $m \in \mathbb{N}$ and some $q_D \in (0,1)$. Then there exists some threshold committee size $\hat{n}$ s.t. for any committee size $n \geq \hat{n}$ and any $p \geq \frac{1}{2}$, the pair (q_H, q_D) with $q_H = \hat{q}_H(q_D)$ implies at least m conflict profiles.*

Proof of Theorem 2.A. We apply the same reasoning and methods as in the proof of Theorem 2 to a hawk holding an i-signal. This leaves us again with a case distinction in terms of y_D. To abbreviate notation, write $\gamma := \frac{2p}{1-p} \geq 2$.

The case $y_D \geq 2$: To guarantee m conflict profiles, fix

$$q_H = \beta\left(x_D + m, y_D - 2m, z_D + m\right) = \frac{\gamma^{y_D - 2m} + \gamma^{z_D + m}}{\gamma^{x_D + m} + \gamma^{y_D - 2m} + \gamma^{z_D + m}},$$

assuming $y_D \geq 2m$.

Similarly to the proof of Theorem 2, summing up incentives from profiles $Piv_1(r), Piv_1^T(r), Piv_3(r)$ for $r \leq m-1$ yields a positive value. All remaining profiles incentivize truthtelling.

The case $y_D < 2$: Fix

$$q_H = \beta\left(x_D + 1, y_D - 1, z_D\right) = \frac{\gamma^{y_D - 1} + \gamma^{z_D}}{\gamma^{x_D + 1} + \gamma^{y_D - 1} + \gamma^{z_D}}$$

with y_D being either 0 or 1. This choice guarantees conflict for any profile of type $Piv_2(s) + (1,0,0)$ with $s \leq \left\lfloor \frac{z_D - y_D + 1}{3} \right\rfloor - 1$, and since n and therefore z_H is arbitrarily large while $y_H < 2$, this provides at least m conflict profiles. Note that no pivotal profiles of type Piv_1, Piv_1^T or Piv_3, Piv_3^T exist (except possibly $Piv_3(0) = Piv_2(0)$, $Piv_3^T(0) = Piv_2^T(0)$).

Summing up incentives from profiles $Piv_2(s), Piv_2^T(s), Piv_4(s)$ for $0 \leq s \leq \left\lfloor \frac{\Delta_D+1}{3} \right\rfloor - 4$ as well as from profiles $Piv_2\left(\left\lfloor \frac{\Delta_D+1}{3} \right\rfloor - t\right), Piv_2^T\left(\left\lfloor \frac{\Delta_D+1}{3} \right\rfloor - t\right)$ for $t = 0, \ldots, 3$ and $Piv_4\left(\left\lfloor \frac{\Delta_D+1}{3} \right\rfloor - 3\right)$ provides positive values for sufficiently large n. All remaining profiles of type $Piv_4(s)$ incentivize truthtelling. $\square$

Proof of Lemma 5. We can consider β_λ as a function of three continuous variables n, n_g, Δ. To prove the claim, it is then sufficient to show that

$$\frac{\partial \beta_\lambda\left(n, n_g, \Delta\right)}{\partial \Delta} \geq 0,$$

with equality if and only if $\lambda = 0$. We have

$$\begin{aligned}
\frac{\partial \beta_\lambda (n, n_g, \Delta)}{\partial \Delta} &\geq 0 \\
\Leftrightarrow \quad \frac{\partial}{\partial \Delta}\left(p_r^{n-n_g} \cdot p_\lambda^{\frac{n_g+\Delta}{2}} \cdot p_{r,\lambda}^{\frac{n_g-\Delta}{2}} + p_r^{n-n_g} \cdot p_{r,\lambda}^{\frac{n_g+\Delta}{2}} \cdot p_\lambda^{\frac{n_g-\Delta}{2}}\right) &\geq 0 \\
\Leftrightarrow \quad \frac{\partial}{\partial \Delta}\left(p_\lambda^{\frac{\Delta}{2}} \cdot p_{r,\lambda}^{-\frac{\Delta}{2}} + \cdot p_{r,\lambda}^{\frac{\Delta}{2}} \cdot p_\lambda^{-\frac{\Delta}{2}}\right) &\geq 0 \\
\Leftrightarrow \quad \ln\left(\frac{p_\lambda}{p_{r,\lambda}}\right) \cdot \left(\left(\frac{p_\lambda}{p_{r,\lambda}}\right)^{\frac{\Delta}{2}} - \left(\frac{p_{r,\lambda}}{p_\lambda}\right)^{\frac{\Delta}{2}}\right) &\geq 0
\end{aligned}$$

and as $\frac{p_\lambda}{p_{r,\lambda}} \geq 1$ with equality if and only if $\lambda = 0$, the latter inequality yields the claim. $\square$

References

[1] D. Austen-Smith and J. S. Banks, "Information aggregation, rationality, and the condorcet jury theorem," *The American Political Science Review*, vol. 90, no. 1, pp. 34–45, 1996.

[2] D. Austen-Smith and T. J. Feddersen, "Deliberation, preference uncertainty and voting rules," *American Political Science Review*, vol. 100, pp. 209–218, 2006.

[3] M. Battaglini, "Multiple referrals and multidimensional cheap talk," *Econometrica*, vol. 70, no. 4, pp. 1379–1401, 2003.

[4] A. Blume, O. Board, and K. Kawamura, "Noisy talk," *Theoretical Economics*, vol. 2, no. 4, pp. 395–440, 2007.

[5] P. J. Coughlan, "In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting," *American Political Science Review*, vol. 94, pp. 375–393, 2000.

[6] V. P. Crawford and J. Sobel, "Strategic information transmission," *Econometrica*, vol. 50, no. 6, pp. 1431–1451, 1982.

[7] D. Dickson, C. Hafer, and D. Landa, "Cognition and strategy: A deliberation experiment," *Journal of Politics*, vol. 70, pp. 974–989, 2008.

[8] J. Elster, "The market and the forum: three varieties of political theory," *Deliberative democracy: Essays on reason and politics*, pp. 3–34, 1997.

[9] T. Feddersen and W. Pesendorfer, "Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting," *The American Political Science Review*, vol. 92, no. 1, pp. 23–35, 1998.

[10] D. Gerardi, R. McLean, and A. Postlewaite, "Aggregation of expert opinions," *Games and Economic Behavior*, vol. 65, no. 2, pp. 339–371, 2009.

[11] D. Gerardi and L. Yariv, "Deliberative voting," *Journal of Economic Theory*, vol. 134, pp. 317–338, 2007.

[12] J. K. Goeree and L. Yariv, "An experimental study of collective deliberation," *Econometrica*, vol. 79, pp. 893–921, 2011.

[13] J. Habermas, *Moral conciousness and communicative action.* MIT press, 1992.

[14] P. Hummel, "Deliberation in large juries with diverse preferences," *Public Choice*, vol. 150, pp. 595–608, 2012.

[15] M. T. Le Quement, "Communication compatible voting rules," *Theory and Decision*, pp. 1–29, October 2012.

[16] M. T. Le Quement and V. Yokeeswaran, "Deliberation, conflict and unanimity." mimeo, University of Bonn, 2012.

[17] B. Manin, "On legitimacy and political deliberation," *Political theory*, vol. 15, no. 3, pp. 338–368, 1987.

[18] A. Meirowitz, "In defense of exclusionary deliberation: Communication and voting with private beliefs and values," *Journal of Theoretical Politics*, vol. 19, pp. 329–360, 2007.

[19] R. Van Weelden, "Deliberation rules and voting," *Quarterly Journal of Political Science*, vol. 3, pp. 83–88, 2008.

[20] A. Wolinsky, "Eliciting information from multiple experts," *Games and Economic Behavior*, vol. 41, no. 1, pp. 141–160, 2002.