Truthtelling in Heterogeneous Committees*

Inga Deimen†, Felix Ketelaar‡, Mark Le Quement§

August 13, 2012

Abstract

This paper analyses truthtelling incentives in pre-vote communication in heterogeneous committees. We modify the classical Condorcet jury model by introducing a new informational structure that captures the feature of 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 frequently an equilibrium outcome of our model. Furthermore, abandonning the assumption of sincere voting, we characterize necessary and sufficient conditions for the existence of truthful equilibria implementing the first best decision rule. These conditions are met for a large set of parameter values implying the possibility of ex post conflict between committee members.

JEL codes: D72, D82, D83

1 Introduction

We consider a situation of collective decision making in which privately informed heterogeneous agents communicate before casting their vote. We show that in our model, heterogeneity and ex post conflict are compatible with full information pooling followed by sincere voting. This follows from the crucial role played by informational consistency in our model. Our possibility result stands in stark contrast to the negative insights yielded by existing literature.

In the classical Condorcet jury model, the defendant is simply either guilty or innocent, and bits of available evidence accordingly either indicate the one or the

*We thank D. Szalay and D. Krähmer for helpful suggestions and encouragement. Furthermore, this paper has benefitted from comments made by the following audiences: Bonn Microeconomic Workshop, SFB conference in Mannheim, Spring Meeting of Young Economists in Mannheim, Annual Congress of the Society of Social Choice and Welfare in New Delhi, EEA Conference in Malaga, Annual Meeting of the Verein für Socialpolitik in Göttingen.

†BGSE, University of Bonn

‡BGSE, University of Bonn

§Institute of Microeconomics, University of Bonn

other of these two states. In practice, there are usually many mutually exclusive modalities according to which the defendant could be guilty or innocent. If he is guilty, he has committed the crime in one of several possible ways. Similarly, if he is innocent, he must have been indulging in some other activity. In other words, each of the basic states of the original model (guilt or innocence) is better understood as a set of substates of the world, each substate representing a separate instance of the basic state that it incarnates.

In our model, a single piece of evidence will always indicate a particular modality of guilt or innocence. Evidence pointing towards the same basic state of the world will exhibit varying degrees of coherence depending on how consistently it indicates the same substate of a given basic state. The more consistently signals indicate a given substate, the stronger the evidence for the basic state that this substate 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. Our first theorem 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, there is a large set of parameter values for which the TS equilibrium exists. Our second theorem states that arbitrary amounts of ex post conflict are compatible with the existence of the TS equilibrium in large committees. This result shows that the number of informational scenarios in 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 the result implies that, for a fixed profile of juror types, increasing committee size ultimately guarantees the existence of the TS equilibrium. In the third part of our analysis, we release the assumption of sincere voting and consider coordinated voting equilibria featuring weakly dominated voting. For general committees, we identify the necessary and sufficient conditions under which the welfare maximizing decision rule can be implemented. We furthermore show that if it is implementable by any mechanism, then it is implementable in a truthful equilibrium of the simple communication and voting game. We show, finally, that the identified implementability conditions are satisfied for a large set of parameter values and independent of the chosen voting rule (excluding unanimous rules).

The difference between our results and the classical findings originates in new strategic effects that arise within our framework. In the classical model, the pivotality of a given juror in the communication stage pins down uniquely the information held by remaining committee members. In the TS equilibrium of our model, this uniqueness breaks down due to our more complex information structure. From the multiplicity of pivotal scenarios featured within our model,

two forces arise which each incentivize truthtelling even if the committee is heterogeneous in the sense that there is a positive probability of ex post conflict among jurors. 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 likelihood of conviction. 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 classical model, at least one juror type will deviate from truthtelling, thereby increasing the likelihood that his favorite decision is taken.

We conclude by highlighting three methodological contributions of this paper. First, our model overturns the impossibility results arising from the classical model on the basis of only a minimal modification of the latter. Second, our model can be interpreted as a generalization of the classical model. We show in Appendix A that the classical model can be nested as a special case in ours. Finally, our analysis constitutes, to the best of our knowledge, the first exploration of the role of consistency in communication problems. The concept of consistency allows to uncover new strategic effects and at the same time captures a fundamental empirical aspect of information structures.

Related literature. A milestone in the theoretical 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 determine a breakdown of the TS equilibrium.

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 contrast, in our environment, sequential communication does not always eliminate the TS equilibrium. 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 heterogenous committees. This latter caveat does not apply to our model, where arbitrary amounts of conflict are 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), despite the presence of heterogeneity. 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 and modifies the equilibrium prediction. 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. A main result of the second contribution is that under unanimity, when subgroup deliberation leads to at least one reactive equilibrium, there exists at least one such equilibrium that is Pareto welfare improving with respect to any symmetric private voting equilibrium. In an extension to our main analysis, we consider an alternative to the full pooling scenario involving partial information sharing. The equilibrium that we study involves coarse public communication as opposed to semi-public communication.

A set of positive and normative contributions stress the importance of the full pooling scenario. The experimental work of Goeree and Yariv (2011) documents extensive truthtelling even 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.

We proceed as follows. Section 2 presents the model. Section 3 develops a simple

formal example identifying key mechanisms. Section 4 presents our main results on the existence of the TS equilibrium. Section 5 relaxes the assumption of sincere voting and considers the possibility to implement the first best decision rule through truthful equilibria. Section 6 considers sequential voting. Section 7 concludes.

2 The model

In this section we present our model in formal terms 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. We assume a uniform prior and split only one of the basic states (G) into only two substates (G_1 and G_2) for the purpose of simplicity. These assumptions suffice to capture the main effects of our model and can easily be generalized.

The jury decides about which action $a \in \{A, C\}$ to implement according to some prespecified voting rule $k \in \{2, \ldots, n-1\}$. We rule out the two types of unanimity voting rules for reasons that we discuss later on. 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.

The utility of juror j depends on the underlying state, the implemented action and on 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, independently of the implemented 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, hawks and doves, 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 5 we extend our results to the case of individual preference parameters.

Prior to the voting stage, each juror j receives a private signal $s_j \in \{i, g_1, g_2\}$. Signals are i.i.d. conditional on the realized state of the world ω. Signals show the correct state of the world with probability $p \in \left(\frac{1}{3}, 1\right)$ while they indicate either of the remaining states with probability $\frac{(1-p)}{2}$.

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 innocent signals, y the total number of g_1-signals and z the total number of g_2-signals held by the committee. In particular, Bayesian posteriors about the probability of the defendant being in one of the two guilty states $\{G_1, G_2\}$ are 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 information structure. Keeping the total number of signals as well as the total number of guilty signals fixed, the posterior probability of guilt is increasing in this difference.

In Appendix A, we relate our environment to a setting where information about which of the two modalities of guilt applies is garbled. In this latter setting, as in the classical model, jurors only learn about the defendant being guilty or innocent. We thereby show that the classical setup can be nested within our general environment.

After having received their signals, jurors communicate through a round of simultaneous and public cheap talk. Given the particular equilibrium that we study later on, 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:

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

Our equilibrium concept is Perfect Bayesian Equilibrium. Yet, for almost the entire paper we are concerned only 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 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.

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 w.r.t. 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 $(0,2,1)$. So there are two signal profiles for which hawks and doves disagree on the optimal decision, namely $(0,1,2)$ and $(0,2,1)$.

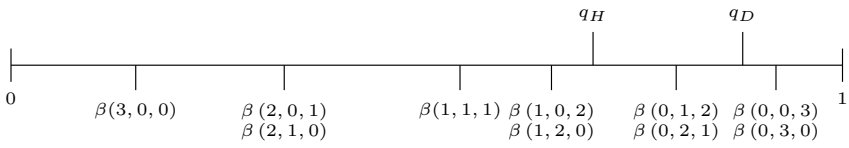

Figure 1: Conditional probabilities of guilt.

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, an individual juror, no matter his preference type, has no 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 favoured decision ensues. In other words, deviating at the voting stage is a dominated strategy. Secondly, given that the voting rule is simple majority, doves are always able to implement their favoured decision. Indeed, if the hawk favours conviction but the doves do not, they can implement acquittal simply by voting for it while if the hawk prefers acquittal, the 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 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 either $(1,2,0)$ 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 favoured 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 conflict profiles, a hawk juror with an i-signal fully agrees with the doves on the preferred action in all pivotal scenarios.

Assume 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 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 after deviation of the hawk is either $(1,0,2)$ or $(1,2,0)$, thus leading to an additional 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 $(0,0,2)$, given other jurors' information, 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, which itself depends on the probability assigned to the states G_1 and G_2. Now, 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 aggregate signal profiles implying disagreement between preference types.

4 The general case

The example of Section 3 yields a simple possibility result, showing that the TS equilibrium can exist despite potential disagreement after full pooling of information. This section extends our analysis to arbitrary committee sizes and general constellations of preference parameters. Section 4.1 introduces key notions. Section 4.2 establishes our main results concerning the existence of the TS equilibrium. Section 4.3 presents a numerical illustration of our results. Proofs are relegated to Appendix B.

4.1 Key notions

Fix the number n of jurors. 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. For any preference type q_j, $j \in \{H, D\}$, unless it prefers acquittal for any possible realization of signals, there exists a unique *threshold profile* (x_j, y_j, z_j), up to transposition of the last two entries, 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. Without loss of generality, we denote threshold profiles in such a way that $z_j \geq y_j$.

As we assume $q_D > q_H$, heterogeneity among jurors is part of our model by construction if there is at least one hawk and one dove. Yet, given the discrete structure of information, heterogeneity does not in itself necessarily imply different preferences over outcomes at some information set. 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 is directly related 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 is 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 favoured decison 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 with critical mass decides according to threshold profile (x_j, y_j, z_j), the defendant will be convicted if and only if the (truthfully) revealed signal profile (x, y, z) satisfies

$$\beta(x_j, y_j, z_j) \leq \beta(x, y, z).$$

4.2 Existence of the TS equilibrium

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 pivotal profiles is required. Appendix A provides an explicit classification of pivotal profiles as well as a description of the relation between the multiplicity of pivotal profiles and the phenomenon of consistency.

For a fixed voting rule, the problem of checking the existence of the TS equilibrium essentially resides in the analysis 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 in a committee featuring ex post conflict within the classical model. We provide a treatment of these deviation incentives in the proofs of our theorems.

In the following preliminary comments, we rule out all remaining deviations. First, note that in the putative TS equilibrium, no juror has an incentive to deviate in the voting stage. This follows trivially from the definition of the TS equilibrium. Secondly, in the putative TS equilibrium, no juror of the preference

type that has critical mass has an incentive to deviate in 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 1. *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).*

Lemma 1 follows from a non-trivial argument that relies on the uncertainty effect described in the example of Section 3. A juror holding a g_2-signal assigns higher probability to scenarios incentivizing truthtelling than to those incentivizing misreporting a g_1-signal.

Given Lemma 1 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, by deviating from truthtelling, to overturn a conviction (an acquittal).

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 $\hat{q}_D(q_H)$, for which an analytical expression is provided in the proof. The TS equilibrium exists iff $q_D \in (q_H, \hat{q}_D(q_H)]$. Part b) yields a fundamental qualitative statement: For virtually any hawk type q_H, in particular for any $q_H \in I_n = (\beta(n-1,0,1), \beta(0,1,n-1))$, there exist dove types q_D such that the TS equilibrium exists despite the fact that they imply at least one conflict profile, given q_H. Note that $I_n \to (0,1)$ as $n \to \infty$. Figure 2, below, provides a graphical illustration of the statement.

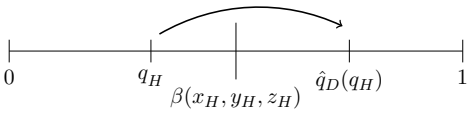

Figure 2: The critical dove type implies conflict.

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.

While the deviation considered in Lemma 1 could be preempted thanks to the uncertainty effect, the deviation considered in the final proof of Theorem 1 (a dove deviating from some g-signal to an i-signal) is disincentivized exclusively by the consensus effect. Deviating from some g-signal to an i-signal systematically increases the chance of an acquittal, but there are in fact many pivotal profiles at which truthtelling implies a conviction that a dove considers optimal. A dove juror, when considering lying about a g-signal, thus trades off the possibility of adversely reverting a conviction against the possibility of beneficially reverting a conviction.

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. Note that the same argument appears in Austen-Smith and Feddersen (2006). Interestingly, in our environment, and in contrast to the latter setting, unanimity can be compatible with the existence of the TS equilibrium, as is the case in the example of Section 3 if assuming $k = 3$ instead of $k = 2$.

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

Theorem 2. *Let hawks have critical mass.*

a) Fix some $m \in \mathbb{N}$ and some $q_H \in (0,1)$. 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 $q_D = \hat{q}_D(q_H)$ implies at least m conflict profiles.

b) Fix some $m \in \mathbb{N}$ and some $\epsilon > 0$. Then there exists some threshold committee size $\hat{\hat{n}}$ s.t. for any committee size $n \geq \hat{\hat{n}}$, any $q_H \in [\epsilon, 1-\epsilon]$ and any $p \geq \frac{1}{2}$, the pair (q_H, q_D) with $q_D = \hat{q}_D(q_H)$ implies at least m conflict profiles.

Both parts of Theorem 2 state 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. Part a) guarantees the existence of the TS

equilibrium for a given number of conflict profiles uniformly over signal precisions and large committee sizes whereas Part b) additionally guarantees the existence of the TS equilibrium simultaneously for most values q_H.

Theorem 2 provides the general insight that the number of conflict profiles, in our model, has no 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 finding stands in fundamental contrast to results in the literature on deliberation in heterogeneous committees; while Coughlan (2000) and Van Weelden (2008) show that one conflict profile already precludes the existence of the TS equilibrium, Le Quement (2012) shows that uncertainty about juror preferences (as in Austen-Smith and Feddersen (2006)) does not suffice to guarantee the existence of the TS equilibrium in large committees if there exists more than a single conflict profile.

Theorem 2 helps us answer whether, for a given pair (q_H, q_D), the TS equilibrium exists for a sufficiently large committee size. This question complements Part a) of Theorem 1 which, for a given q_H, identifies a critical dove type $\hat{q}_D(q_H)$ for fixed n. For a given committee of size n the question is whether the existence of the TS equilibrium may be ensured by simply increasing the number of jurors while maintaining the original juror types as well as the allocation of critical mass. In what follows, we establish a simple sufficient condition ensuring that this is indeed the case.

For any pair (q_H, q_D), Theorem 2 guarantees that the TS equilibrium exists for a sufficiently large committee if the maximal number of conflict profiles implied by the pair (q_H, q_D), across all committee sizes, is bounded. A necessary and sufficient condition guaranteeing this boundedness is that $q_H,\ q_D \in \left(\frac{\left(\frac{2p}{1-p}\right)^r}{1+\left(\frac{2p}{1-p}\right)^r}, \frac{\left(\frac{2p}{1-p}\right)^{r+1}}{1+\left(\frac{2p}{1-p}\right)^{r+1}}\right]$, for some $r \in \mathbb{Z}$ (cf. Lemma B.1 in Appendix B). Note that this interval condition on (q_H, q_D) is reminiscent of the main condition stated in Coughlan (2000). In his model, for any even committee size, a TS equilibrium exists iff $q_H, q_D \in \left(\frac{\left(\frac{p}{1-p}\right)^{2r}}{1+\left(\frac{p}{1-p}\right)^{2r}}, \frac{\left(\frac{p}{1-p}\right)^{2(r+1)}}{1+\left(\frac{p}{1-p}\right)^{2(r+1)}}\right]$, for some $r \in \mathbb{Z}$. While the interval condition provided in Coughlan (2000) is both necessary and sufficient as well as independent of committee size, our interval condition is simply sufficient (and not necessary) to guarantee the existence of the TS equilibrium in large committees.

We now give an intuition for the fact that increased committee size can be helpful in establishing the TS equilibrium. Consider a pair (q_H, q_D) such that the maximal number of conflict profiles implied, across all committee sizes, is bounded. Consequently, for such a pair, as committee size increases, the number of pivotal profiles implying an incentive to lie is bounded as well. At the same time, as committee size increases, the total number of pivotal profiles tends to infinity, as shown in Lemma B.1 of Appendix B. The relative share of pivotal profiles implying an incentive to tell the truth thus tends to one as committee size tends to infinity, so that the trade off between lying and truthtelling ultimately

resolves in favour of truthtelling. This positive effect of increased committee size stands in stark contrast to the adverse effect documented in Meirowitz (2007) and Le Quement (2012).

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 Appendix B as Theorems B.1 and B.2.

4.3 Numerical illustrations

To close this section, we complement our analytical results concerning the existence of the TS equilibrium with some numerical and graphical illustrations that highlight various features of our model for the classical jury size $n = 12$.

The two step functions in Figure 3 indicate, for every value of q_H, the maximal dove type s.t. the TS equilibrium exists, for a given type holding critical mass. Fixing $p = 0.8$, the solid step function corresponds to the case where hawks have critical mass, while the dashed step function corresponds to the case where doves have critical mass. Given that by definition, $q_H < q_D$, the feasible parameter pairs are located in the area above the dotted 45°-line.

The two most relevant insights of Figure 3 are as follows: First, the set of pairs (q_H, q_D) for which the TS equilibrium exists under some voting rule is large, covering more than half of the area above the 45°-line. Secondly, note that there is a substantial set of pairs (q_H, q_D) for which the existence of the TS equilibrium depends on the chosen voting rule. Note that in particular, no voting rule unilaterally outperforms the other 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).

Figure 4 focuses on the aspect of conflict between jurors. Keeping $p = 0.8$ fixed, for each value of q_H, it indicates the maximum number of conflict profiles compatible with the TS equilibrium given an optimal choice of the voting rule. As discussed before, the existence of the TS equilibrium in Coughlan's model depends on the (non-)existence of conflict profiles. Figure 4 clearly indicates that under a standard signal precision, the feasible number of conflict profiles in our model is substantially larger than one for any reasonable value of q_H.

Figure 5, finally, explores the sensitivity of the TS equilibrium to the signal precision p. Focusing on the case where hawks have critical mass, the critical dove types $\hat{q}_D(q_H)$ are plotted as a function of q_H for $p = 0.5$ (dot-dashed), $p = 0.65$ (dashed) and $p = 0.8$ (solid). Typically, $\hat{q}_D(q_H)$ is seen to increase in p, so Figure 5 indicates a clear tendency for an increased signal precision to ease the parametrical conditions on (q_H, q_D) guaranteeing the existence of the TS equilibrium. Yet, the effect of p is not monotonic in general, as appears clearly in the lower left corner of the figure, where the solid step function crosses the dashed step function from above.

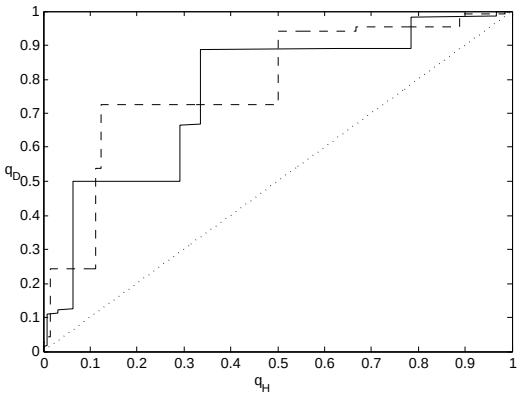

Figure 3: Preference types compatible with TS equilibrium.

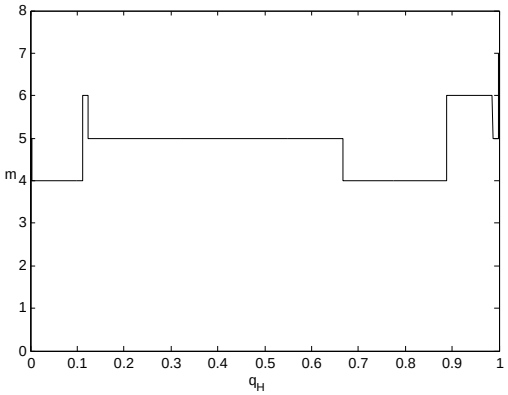

Figure 4: Number of conflict profiles compatible with TS equilibrium.

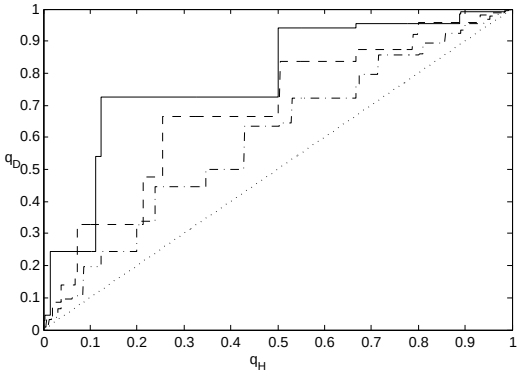

Figure 5: Preference types compatible with TS equilibrium for different signal precisions.

5 Optimal committee design

So far, we have restricted our analysis to the study of equilibria involving truthful deliberation followed by sincere voting. We now expand the set of equilibria by considering the possibility of weakly dominated voting. If voting strategies are such that given the observed profile of reports, a voter is not pivotal in the voting stage, a juror has no strict incentive to vote for his favoured outcome. Once using the publicly observed report profile as a device for the coordination of votes, truthtelling can be ensured as an equilibrium outcome in more general settings than previously considered.

General committees. Throughout the previous sections, we assumed that the committee consists of two homogeneous subgroups described as respectively hawks and doves. A natural generalization of this basic 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 q the rule that assigns to each profile of pooled signals the decision that a juror of preference type q favours. We now state a result pertaining to the existence of an equilibrium that implements a decision rule q. 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 q would favour a conviction. Accordingly, an individual juror is never pivotal in the voting stage as long as the voting rule is non-unanimous.

The following theorem makes use of results obtained in the theorems 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 q if and only if $q_1 \geq \hat{q}_H(q)$ and $q_n \leq \hat{q}_D(q)$.*

Theorem 3 states that a truthful equilibrium implementing decision rule q exists iff the thresholds of the most extreme jurors q_1 and q_n are within the interval $[\hat{q}_H(q), \hat{q}_D(q)]$ defined by the threshold of a virtual juror of preference type q. Note that, as compared to the 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.

We visualize Theorem 3 in Figure 6 for $n = 12$ and $p = 0.8$. Given any decision rule 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 q. For every q, a truthful equilibrium implementing decision rule q exists if and only if all juror types are located within the interval defined by the two step functions.

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 q. This result is reminiscent of Gerardi and Yariv (2007).

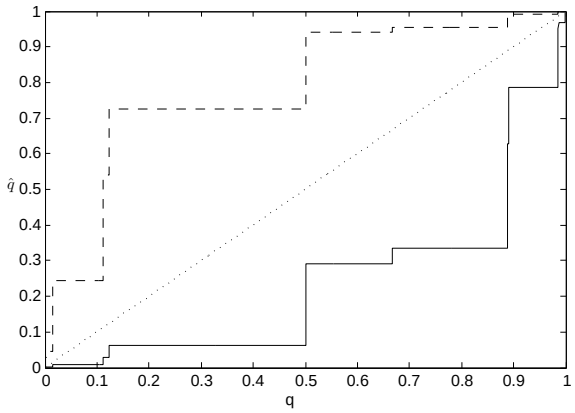

Figure 6: Preference types compatible with truthful equilibrium implementing decision rule q.

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

$$Eu = E\sum_{j=1}^{n} \lambda_j u_j = \sum_{j=1}^{n} \lambda_j Eu_j$$

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

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

so

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

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

Proposition 1. *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.*

Proposition 1 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 favour 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. An attractive feature of Proposition 1 is that the welfare maximizing decision rule can be implemented without resorting to an outside agent endowed with commitment power.

In what follows, we analyze the possibility of implementing the social welfare maximizing decision rule in general committees through truthful revelation and sincere voting. We examine two types of settings. The first involves the simple communication and voting procedure analyzed in section 4. The second involves the use of a mediator that centralizes private reports and subsequently emits recommendations.

Suppose the decision rule is given by k. The conditions of Theorem 3 do not necessarily suffice to guarantee the existence of a TS equilibrium that implements decision rule q_k. Indeed, the problem is reminiscent of the case of a two types committee under unanimity. Consider some juror that is strictly harsher than juror k and hence has a potential incentive to misreport an i-signal as some g-signal. Suppose he does so, and suppose furthermore that after the communication stage, a conviction is the preferred outcome of exactly jurors 1 to k if everyone bases his decision on the reported signals only. If the juror who deviated at the communication stage now becomes aware that, given the actual signal distribution, he prefers an acquittal, he can implement this outcome by voting for acquittal. The same reasoning holds for more lenient juror compared to juror k in an equivalent fashion. By the above argument, if jurors $k-1$, k and $k+1$ agree on the optimal decisions for all possible signal realizations, the

conditions of Theorem 3 imply the existence of the TS equilibrium. While this is clearly not satisfied in a general committee, it can be ensured by a designer who aims to optimize the composition of the committee.

The alternative approach of installing a mediator of preference type q who 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.

6 Sequential communication

Under sequential communication, the TS scenario is often not an equilibrium although it constitutes an equilibrium under simultaneous communication. Van Weelden (2008) shows that under sequential communication, even under uncertainty about preference types, the TS equilibrium does not exist, in contrast to the simultaneous case (see Austen-Smith and Feddersen (2006)), unless there is guaranteed to be no ex post conflict in the committee.

The main drawback of sequential communication is that it reduces uncertainty for late speakers. Suppose a sequential communication protocol in our original setup with only two preference types. Consider the final speaker in the putative sequential TS equilibrium: Given the public reports of previous speakers, the final speaker knows exactly the consequences of his report. In particular, suppose the following scenario. First, previously announced signals constitute a pivotal signal profile. Secondly, previous signals are such that, together with the remaining juror's signal, the overall profile is a conflict profile. Then, the last juror necessarily has access to a strictly advantageous deviation if he is not part of the group having critical mass.

In contrast to Van Weeldens's result, in our setup, the TS equilibrium may exist under sequential communication. To demonstrate this, revisit the example of Section 3 involving two doves and one hawk. As shown there, the TS equilibrium exists under simultaneous communication for $k = 2$. Now, consider a sequential protocol where the hawk speaks first while doves report afterwards. Clearly, having critical mass, the doves have no incentive to misreport. The hawk, when speaking first, faces identical incentives as in the TS equilibrium of the simultaneous communication game. Indeed, when reporting, he knows nothing about the other jurors' information. The TS equilibrium therefore exists under this reporting order.

The above example immediately generalizes in the following way:

Proposition 2. *Suppose a jury consisting of hawks and doves and suppose the TS equilibrium exists under simultaneous communication if hawks (doves) have*

critical mass. Then, if there is only one dove (hawk), the TS equilibrium also exists under sequential communication, if the single juror belonging to the group not having critical mass reports first.

Note that this result typically does not extend if the number of jurors in the group not having critical mass is larger than one. Indeed, consider a variant of the example of Section 3 with two hawks, one dove and unanimity $k = 3$, so that doves still have critical mass. It is easy to check that the TS equilibrium in this context still exists even though the voting rule is now unanimity. As discussed in Section 3, the only pivotal profiles faced by a hawk are $(0, 2, 0)$ and $(0, 0, 2)$. In the putative sequential TS equilibrium, at least one hawk reports after another juror. Suppose now that one previous report revealed a g_1-signal, so that the only pivotal profile faced by the second hawk speaker is $(0, 2, 0)$. If the latter hawk holds a g_2-signal, he prefers conviction at this pivotal signal profile. However, given the dove's preferences, the only way to implement conviction is to deviate from truthtelling by reporting a g_1-signal instead of a g_2-signal.

Partially revealing equilibria and welfare. We conclude this section by examining partial pooling equilibria with sincere voting.

In the standard binary signal model, equilibria involving imperfect individual communication must involve mixed reporting strategies. In our setup, an intuitive alternative to the TS equilibrium is an equilibrium in which some jurors adopt the following strategy: they report "My signal indicates guilt." when holding a g_1- or a g_2-signal, without further specifying the modality of guilt. We demonstrate by means of an explicit example that such an equilibrium can exist and may furthermore improve overall committee welfare as compared to the TS equilibrium. Interestingly, in our example, this equilibrium arises more naturally under a sequential communication protocol than a simultaneous one.

Fix $n = 3$, $k = 3$, and let the jury contain two hawks and one dove. Furthermore, assume $p = \frac{1}{2}$, and $q_D = \frac{7}{8}$, $q_H = \frac{3}{4}$. Clearly, we have

$$\begin{aligned} \beta(1,0,2) &= \frac{5}{7} < q_H < \frac{6}{7} = \beta(0,1,2), \\ \beta(0,1,2) &= \frac{6}{7} < q_D < \frac{9}{10} = \beta(0,0,3). \end{aligned}$$

Preferences are thus as in Section 3 and the TS equilibrium exists if $k = 3$, i.e. if doves have critical mass. Yet, attaching equal Pareto weights $\lambda_j = \frac{1}{3}$ to all jurors, the socially optimal decision rule is to implement conviction for signal profiles $(0, 2, 1)$, $(0, 1, 2)$, $(0, 3, 0)$, $(0, 0, 3)$, i.e. it coincides with the hawks' optimal decision rule. Unfortunately, if hawks have critical mass, i.e. $k \leq 2$, the TS equilibrium does not exist as the dove, when holding a g-signal, strictly prefers to deviate towards reporting i, as can readily be checked. So the TS equilibrium exists only under unanimity and hence comes at the cost of suboptimally acquitting for profiles $(0, 2, 1)$ and $(0, 1, 2)$.

In contrast, consider a sequential reporting game while keeping unanimity as the voting rule. Assume that jurors 1 and 2 are hawks while juror 3 is a dove and jurors report in the order determined by their subscript. Consider the following putative partial revelation equilibrium. Whenever juror 2 holds a g-signal and juror 1 reported either g_1 or g_2, juror 2 reports the same modality of guilt as juror 1, whether or not his signal indeed matches the modality reported by juror 1. In remaining cases, juror 2 reports truthfully. Finally, jurors vote sincerely at the voting stage.

We now examine outcomes in this putative partial revelation equilibrium. Clearly, if some juror holds an i-signal, this signal will be truthfully revealed and the (consensual) outcome is acquittal. If all jurors hold some signal indicating guilt the situation is as follows: Whenever jurors 1 and 3 hold inconsistent g-signals, the defendant is acquitted. Indeed, as juror 1 truthfully reports his signal, juror 3, being a dove, trivially prefers acquittal independently of the signal of juror 2. The critical case is when jurors 1 and 3 hold signals which indicate identical modalities of guilt. Suppose jurors 1 and 3 both hold a g_2-signal. Juror 2, despite having reported a g_2-signal, might still hold a g_1-signal given that his report was not necessarily truthful. So the true aggregate signal profile can be either $(0, 1, 2)$ or $(0, 0, 3)$ and juror 3 faces uncertainty about which outcome to prefer. Computing utilities from conviction resp. acquittal under these beliefs yields a higher expected utility of conviction, hence juror 3 will implement conviction. Our putative partial revelation equilibrium is thus an equilibrium of the game. Moreover, committee welfare in this equilibrium is strictly improved as compared to the TS equilibrium under unanimity that exists under simultaneous communication. Indeed, in addition to signal profiles $(0, 3, 0)$ and $(0, 0, 3)$, the defendant will also be convicted for some signal profiles of type $(0, 2, 1)$ and $(0, 1, 2)$, namely for signal distributions $s_1 = g_1$, $s_2 = g_2$, $s_3 = g_1$ and $s_1 = g_2$, $s_2 = g_1$, $s_3 = g_2$.

Note that the above partial revelation outcome can be replicated under a simultaneous communication protocol in an equilbrium where one hawk and one dove reveal their signals truthfully while the remaining hawk reveals an innocent signal but always reports g_1 if holding a g_1- or a g_2-signal. However, this equilibrium requires that ex ante identical jurors behave differently, which may appear unnatural. In the sequential setup the ex ante symmetry is intrinsically broken by the communication protocol.

The above analysis demonstrates that full pooling of information at the communication stage can be welfare dominated by an equilibrium involving only partial revelation. Interestingly, in this light, sequential communication can serve as a tool for equilibrium selection. A sequential communication protocol potentially increases committee welfare by ruling out the natural but welfare dominated TS equilibrium that might exist under simultaneous communication.

7 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.

Our analysis points to the challenge of identifying a good indicator of the difficulty of full information pooling in heterogeneous committees. Clearly, as revealed by Theorem 2, the number of signal profiles for which ex post conflict arises (as used in Austen-Smith and Feddersen (2005), Le Quement (2012)) is not an adequate predictor of the existence of the TS equilibrium in our model.

On the other hand, heterogeneity of preference types is arguably significantly more informative in our model than in the classical model considered in Coughlan (2000). In the latter, there are always reasonable preference type constellations for which the TS equilibrium does not exist despite preference types being arbitrarily close (but not identical). In contrast, in our model, a moderate spread in preference types is always compatible with the possibility of full information pooling unless juror types are extreme.

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 setup, a juror, when choosing his announcement, cannot rank g-reports in terms of how much evidence of guilt each offers. This will depend on the consistency of each g-report with other jurors' reports. In other words, in our environment, announcements exhibit a form of complementarity. We believe that the above instance of contextual generation of meaning could fruitfully be studied within the context of other communication games, whether featuring cheap talk or verifiable information.

Appendix A

The main goal of Appendix A is to relate and contrast our information structure to a benchmark information structure in which signals only contain information about whether the defendant is innocent or guilty, as in Coughlan (2000). In particular, we highlight the role played by consistency in our information structure, which is entirely absent from the benchmark information structure.

First, recall the information structure that we assume: 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 they indicate either of the remaining states with probability $p_r = \frac{(1-p)}{2}$, the subscript r standing for "residual". Call this information stucture ungarbled.

Secondly, 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. The signal generating process now takes 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 $\frac{p_r+p_r}{2} = p_r$. However, if, say, G_1 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 $\frac{p+p_r}{2}$.

Thirdly, consider a mixture of the two above introduced information structures parameterized by $\alpha \in [0, 1]$. An α-mixed signal generating process attaches weight α to the ungarbled signal generating process and weight $1 - \alpha$ to the garbled signal generating process. In formal terms, this yields the following:

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 $\alpha \cdot p_r + (1 - \alpha) \cdot \frac{p_r+p_r}{2} = p_r$. Note that none of these numbers depends on α.

If instead 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 $p_\alpha := \alpha \cdot p + (1 - \alpha) \cdot \frac{p+p_r}{2}$ and a g_2-signal is generated with probability $p_{r,\alpha} := \alpha \cdot p_r + (1 - \alpha) \cdot \frac{p+p_r}{2}$. Note that the probabilites of different signals now clearly depend on α. The case of G_2 being the true state of the world is analogous.

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

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

Again, $\beta_\alpha(x, y, z)$ is symmetric with respect to the last two entries, reflecting the symmetric treatment of the two modalities of guilt in our model. Moreover, applying a "change of coordinates", note that the value $\beta_\alpha(x, y, z)$ is fully determined also by the following three numbers: The committee size $n = x + y + z$, the total number of guilty signals $n_g = y + z$ and the absolute value of the difference between the two numbers of guilty-type signals, $\Delta = |z - y|$, where n_g and Δ are either both odd or both even and $n \geq n_g \geq \Delta$. Explicitely, we have

$$\beta_\alpha(x, y, z) = \frac{p_r^{n-n_g} \cdot p_\alpha^{\frac{n_g+\Delta}{2}} \cdot p_{r,\alpha}^{\frac{n_g-\Delta}{2}} + p_r^{n-n_g} \cdot p_{r,\alpha}^{\frac{n_g+\Delta}{2}} \cdot p_\alpha^{\frac{n_g-\Delta}{2}}}{p^{n-n_g} \cdot p_r^{n_g} + p_r^{n-n_g} \cdot p_\alpha^{\frac{n_g+\Delta}{2}} \cdot p_{r,\alpha}^{\frac{n_g-\Delta}{2}} + p_r^{n-n_g} \cdot p_{r,\alpha}^{\frac{n_g+\Delta}{2}} \cdot p_\alpha^{\frac{n_g-\Delta}{2}}}.$$

The number $\Delta = |z - y|$ can thus be interpreted as a measure of consistency among signals indicating guilt: The higher Δ, the more unambiguously signals indicate one of the two modalities of guilt as being the true state of the world. We may now state the following lemma:

Lemma A.1. Consider β_α as a function of n, n_g, Δ as just described. Then, keeping n and n_g fixed, β_α is weakly increasing in Δ and strictly increasing in Δ if and only if $\alpha > 0$.

Proof. See Appendix B.

Lemma A.1 states that, unless signals do not transmit *any* information about the modality of guilt ($\alpha = 0$), increasing consistency among a given total number of guilty signals increases the posterior probability of guilt. This fundamental property of our information structure is key to our results, being a driving force behind the possibility of truthful information exchange between heterogeneous jurors. Note that while for the sake of expositional simplicity, our analysis is restricted to the case $\alpha = 1$, this property is a feature of the model for any $\alpha > 0$, thus suggesting that our results qualitatively carry over to the general case of $\alpha > 0$.

The case of $\alpha = 0$. This case corresponds to the information structure assumed in Coughlan (2000). In what follows, we briefly examine the key elements of this special case and recall why it is never compatible with the existence of the TS equilibrium. Note that if $\alpha = 0$, it follows that $p_0 = p_{r,0} = \frac{p+p_r}{2}$ and β_0, after some algebra, simplifies to

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

In particular, in line with Lemma A.1, β_0 does not depend on Δ. The model is now mathematically equivalent to the classical two states and two signals model with the following specification. The defendant is guilty with prior probability $\frac{2}{3}$. If the defendant is guilty, some g-signal (either g_1 or g_2) is generated with probability $p + \frac{1-p}{2}$ 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 some g-signal is generated with probability $\frac{1-p}{2}$.

The TS equilibrium cannot exist in this setting by the following argument. Suppose that the number of hawks in the committee is at least k, where k is the voting rule, thus implying that the defendant is convicted in a TS equilibrium iff hawks favour a conviction. Note first that for each juror type j, there is a minimal total number n_g^j such that it favours a conviction iff $n_g \geq n_g^j$. Assume furthermore that $q_H,\ q_D$ are s.t. $n_g^D > n_g^H$. We now focus on the truthtelling

incentives of a dove holding some g-signal in a putative TS equilibrium. Such a juror acts as if knowing that his announcement is pivotal, which is the case iff an announced i-signal leads to an acquittal while an announced g-signal (whether g_1 or g_2) leads to a conviction. Given the nature of β_0, this implies that the total number of g-signals held by remaining jurors is $n_g^H - 1$. Consider the following two facts. First, given that the total number of g-signals (including his own) amounts to n_g^H, the dove juror favours acquittal. Secondly, by deviating to announcing an i-signal instead of truthfully announcing his g-signal, the dove can indeed trigger an acquittal. He will thus deviate to announcing an i-signal, which proves that the TS equilibrium does not exist.

The case of $\alpha = 1$. We now refine our analysis of the signal structure by comparing the relative impact of the total quantity and the consistency of available g-signals. In this case, 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 $p > \frac{1}{3}$, it follows that $\frac{2p}{1-p} > 1$ so that $\beta(x, y, z)$ is decreasing in x and increasing in y and z. This captures the 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.

We typically take the number of jurors n and thereby the total number of signals $x + y + z = n$ as given. We are interested in the effect of shifting mass from one entry of β to another. Indeed, substracting one unit from a given entry of β while adding one unit to another replicates the change in beliefs of others achievable by an individual juror misreporting his signal in a TS equilibrium. Understanding the effect of such a switch is thus crucial to understanding deviation incentives in the communication stage.

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 an innocent signal as either guilty signal (and vice versa) in a putative TS equilibrium increases (decreases) the posterior beliefs of other jurors about the probability of guilt and thus unilaterally increases the likelihood of conviction (acquittal).

On the other hand, we know from Lemma A.1 that shifting mass across the y- and z-entries (the two modalities of guilt) increases the posterior probability of guilt if and only if this shift increases the consistency Δ. This implies that two separate aspects need to be considered simultaneously: a larger total number n_g of g-signals may be offset by a lack of consistency among g-signals, when comparing two signal profiles in terms of the implied posterior probability of guilt. The following lemma provides a quantitative comparison of these two aspects.

Lemma A.2. *Let* $(x,y,z),(\tilde{x},\tilde{y},\tilde{z}) \in \mathbb{N}^3$ *with* $(x,y,z) \neq (\tilde{x},\tilde{y},\tilde{z}) \neq (x,z,y)$ *and* $x+y+z = \tilde{x}+\tilde{y}+\tilde{z} = n$. *Let* $n_g = y+z$, $\Delta = |z-y|$ *resp.* $\tilde{n}_g = \tilde{y}+\tilde{z}$, $\tilde{\Delta} = |\tilde{z}-\tilde{y}|$ *such that* $n_g \geq \tilde{n}_g$. *Then*

$$\tilde{\Delta} - \Delta \leq 3 \cdot (n_g - \tilde{n}_g) \quad \Rightarrow \quad \beta(x,y,z) > \beta(\tilde{x},\tilde{y},\tilde{z}).$$

Moreover, if $p \geq \frac{1}{2}$, *then*

$$\tilde{\Delta} - \Delta \leq 3 \cdot (n_g - \tilde{n}_g) \quad \Leftrightarrow \quad \beta(x,y,z) > \beta(\tilde{x},\tilde{y},\tilde{z}).$$

Proof. See Appendix B.

As appears from the second part of Lemma A.2, under moderate assumptions on the signal precision, we can thus easily compare any two signal profiles of same cardinality w.r.t. the posterior probability of guilt that they induce. The result is fundamental to our analysis as it allows us to address the following key question: In a putative TS equilibrium, given the threshold profile of the critical mass type, what are the pivotal signal profiles faced by a juror of the type that does not have critical mass? The answer to this question is a direct corollary of Lemma A.2 and is given explicitly in the following lemma.

Lemma A.3. *Assume* $p \geq \frac{1}{2}$ *and let hawks have critical mass*[1] *with threshold profile* (x_H, y_H, z_H). *Recall that* $\Delta_H = z_H - y_H \geq 0$. *Then the set PIV of all pivotal profiles consists of the following profiles:*[2]

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

[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 that is smaller or equal than w.

where

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

Furthermore, pivotal profiles are ordered as follows:

$$\begin{array}{rcl}
\begin{array}{r}\beta(Piv_4(0)) \\ \beta\left(Piv_4^T(0)\right)\end{array} & < \ldots < & \begin{array}{l}\beta\left(Piv_4(\min\left\{\left\lfloor\frac{\Delta_H-1}{3}\right\rfloor, x_H\right\})\right) \\ \beta\left(Piv_4^T(\min\left\{\left\lfloor\frac{\Delta_H-1}{3}\right\rfloor, x_H\right\})\right)\end{array} \\
< \begin{array}{r}\beta\left(Piv_3(\lfloor\frac{y_H-1}{2}\rfloor)\right) \\ \beta\left(Piv_3^T(\lfloor\frac{y_H-1}{2}\rfloor)\right)\end{array} & < \ldots < & \begin{array}{l}\beta(Piv_3(0)) \\ \beta\left(Piv_3^T(0)\right)\end{array} \\
= \begin{array}{r}\beta(Piv_2(0)) \\ \beta\left(Piv_2^T(0)\right)\end{array} & < \ldots < & \begin{array}{l}\beta\left(Piv_2(\min\left\{\left\lfloor\frac{\Delta_H+1}{3}\right\rfloor, x_H\right\})\right) \\ \beta\left(Piv_2^T(\min\left\{\left\lfloor\frac{\Delta_H+1}{3}\right\rfloor, x_H\right\})\right)\end{array} \\
< \begin{array}{r}\beta\left(Piv_1(\lfloor\frac{y_H}{2}\rfloor-1)\right) \\ \beta\left(Piv_1^T(\lfloor\frac{y_H}{2}\rfloor-1)\right)\end{array} & < \ldots < & \begin{array}{l}\beta(Piv_1(0)) \\ \beta\left(Piv_1^T(0)\right)\end{array}.
\end{array}$$

Proof. See Appendix B.

Appendix B

Proof of Lemma A.1. We can consider β_α as a function of three continuous variables n, n_g, Δ. To prove the claim, it is then sufficient to show that

$$\frac{\partial\beta_\alpha(n,n_g,\Delta)}{\partial\Delta} \geq 0,$$

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

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

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

Proof of Lemma A.2. Recall that

$$\begin{aligned}\beta(x,y,z) &= \frac{\left(\frac{2p}{1-p}\right)^{y}+\left(\frac{2p}{1-p}\right)^{z}}{\left(\frac{2p}{1-p}\right)^{x}+\left(\frac{2p}{1-p}\right)^{y}+\left(\frac{2p}{1-p}\right)^{z}} \\ &= \frac{\left(\frac{2p}{1-p}\right)^{-\frac{\Delta}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\Delta}{2}}}{\left(\frac{2p}{1-p}\right)^{n-\frac{3}{2}n_g}+\left(\frac{2p}{1-p}\right)^{-\frac{\Delta}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\Delta}{2}}}.\end{aligned}$$

We therefore have

$$\begin{aligned}&\beta(x,y,z) > \beta(\tilde{x},\tilde{y},\tilde{z}) \\ \Leftrightarrow\quad &\frac{\left(\frac{2p}{1-p}\right)^{-\frac{\Delta}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\Delta}{2}}}{\left(\frac{2p}{1-p}\right)^{n-\frac{3}{2}n_g}+\left(\frac{2p}{1-p}\right)^{-\frac{\Delta}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\Delta}{2}}} > \frac{\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}}}{\left(\frac{2p}{1-p}\right)^{n-\frac{3}{2}\tilde{n}_g}+\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}}} \\ \Leftrightarrow\quad &\left(\frac{2p}{1-p}\right)^{\frac{3}{2}(n_g-\tilde{n}_g)}\cdot\left(\left(\frac{2p}{1-p}\right)^{-\frac{\Delta}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\Delta}{2}}\right) > \left(\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}}\right).\end{aligned}$$

Clearly, the LHS is increasing in Δ as $\left(\frac{2p}{1-p}\right) > 1$ for any $p > \frac{1}{3}$. To prove the first claim, it therefore suffices to insert $\Delta = \tilde{\Delta} - 3\cdot(n_g - \tilde{n}_g)$:

$$\begin{aligned}&\left(\frac{2p}{1-p}\right)^{\frac{3}{2}(n_g-\tilde{n}_g)}\cdot\left(\left(\frac{2p}{1-p}\right)^{-\frac{\Delta}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\Delta}{2}}\right) \\ =\ &\left(\frac{2p}{1-p}\right)^{\frac{3}{2}(n_g-\tilde{n}_g)}\cdot\left(\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}-3\cdot(n_g-\tilde{n}_g)}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}-3\cdot(n_g-\tilde{n}_g)}{2}}\right) \\ >\ &\left(\frac{2p}{1-p}\right)^{\frac{3}{2}(n_g-\tilde{n}_g)}\cdot\left(\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}+3\cdot(n_g-\tilde{n}_g)}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}-3\cdot(n_g-\tilde{n}_g)}{2}}\right) \\ =\ &\left(\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}}\right).\end{aligned}$$

Note that the inequality is strict indeed as the assumption $\Delta = \tilde{\Delta} - 3\cdot(n_g - \tilde{n}_g)$ together with $(x,y,z) \neq (\tilde{x},\tilde{y},\tilde{z}) \neq (x,z,y)$ and $n_g - \tilde{n}_g \geq 0$ implies $n_g - \tilde{n}_g > 0$.

To prove the second part, recall that Δ resp. $\tilde{\Delta}$ are odd (even) if and only if n_g resp. $\tilde{n}_g$ are odd (even). It therefore suffices to prove the reversed direction for $\Delta = \tilde{\Delta} - 3\cdot(n_g - \tilde{n}_g) - 2$:

$$\begin{aligned}&\left(\frac{2p}{1-p}\right)^{\frac{3}{2}(n_g-\tilde{n}_g)}\cdot\left(\left(\frac{2p}{1-p}\right)^{-\frac{\Delta}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\Delta}{2}}\right) \\ =\ &\left(\frac{2p}{1-p}\right)^{\frac{3}{2}(n_g-\tilde{n}_g)}\cdot\left(\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}-3\cdot(n_g-\tilde{n}_g)-2}{2}}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}-3\cdot(n_g-\tilde{n}_g)-2}{2}}\right) \\ =\ &\left(\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}}{2}+3\cdot(n_g-\tilde{n}_g)+1}+\left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}-1}\right).\end{aligned}$$

Note that $\Delta = \tilde{\Delta} - 3 \cdot (n_g - \tilde{n}_g) - 2$ implies $\tilde{\Delta} - 3 \cdot (n_g - \tilde{n}_g) - 2 \geq 0$ and hence $-\frac{\tilde{\Delta}}{2} + 3 \cdot (n_g - \tilde{n}_g) + 1 \leq \frac{3}{2} \cdot (n_g - \tilde{n}_g) \leq \frac{\tilde{\Delta}}{2} - 1$. So for any $p \geq \frac{1}{2}$ the second part follows from

$$\begin{aligned}
& \left(\frac{2p}{1-p}\right)^{\frac{3}{2}(n_g - \tilde{n}_g)} \cdot \left(\left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}}{2}} + \left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}}\right) \\
\leq \quad & 2 \cdot \left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2} - 1} \\
\leq \quad & \left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}} \\
< \quad & \left(\frac{2p}{1-p}\right)^{-\frac{\tilde{\Delta}}{2}} + \left(\frac{2p}{1-p}\right)^{\frac{\tilde{\Delta}}{2}}.
\end{aligned}$$

Proof of Lemma A.3. 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) 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 A.2, this implies

$$\begin{cases} (y_H - y) \geq 2(z - z_H) > (y_H - y) - 2 & \text{if } y + z \leq y_H + z_H - 1 \\ (y_H - y) > 2(z - z_H) \geq (y_H - y) - 2 & \text{if } y + z > y_H + z_H - 1 \end{cases},$$

the first case yielding precisely the profiles of type Piv_1, Piv_3 as well as $Piv_4(0)$ while the second case yields precisely profiles of type Piv_2 and Piv_4 except for $Piv_4(0)$. The transposed profiles are derived for the case $y \leq z$ in exactly the same way. The ordering of all pivotal profiles is a direct consequence of Lemma A.2.

Proof of Lemma 1. 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 there is no incentive to deviate from reporting a g_2-signal to reporting a g_1-signal. In particular, this is the case if $y = z$. Otherwise $y < z$ and there exists another profile (x, z, y), namely the transposed variant of (x, y, z). Note that a g_2-report will lead to conviction for profile (x, y, z) and to acquittal for profile (x, z, y), and vice versa for a g_1-report. 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:

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

Hence, for any pair of transposed profiles, the juror is either indifferent between lying a reporting truthfully or has a strict incentive to report truthtfully. 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.

Proof of Theorem 1. Without loss of generality, throughout the whole proof 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 an additional g_2-report leads to conviction. Note that PIV_{g_2} depends on q_H resp. (x_H, y_H, z_H).

a) The TS equilibrium exists iff

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

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

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

Equality in the above expression yields $\hat{q}_D(q_H)$.

b) $Eu(g_2) - Eu(i)$ is continuous and linear in q_D as seen above. Assume $q_D = \beta(x_H, y_H, z_H)$ for the moment. Then, doves are indifferent between both outcomes for profile $Piv_4(0) = (x_H, y_H, z_H - 1)$ while for all other pivotal profiles they weakly prefer truthtelling over deviating (cf. Lemma A.3). To prove strict preference for truthtelling we distinguish three cases: If $(x_H, y_H, z_H) = (n, 0, 0)$, hawks will implement conviction independently of the reports and hence the TS

equilibrium trivially exists. If $y_H > 0$, the remaining jurors hold signal profile $Piv_3(0) = (x_H, y_H - 1, z_H)$ with positive probability. From Lemma A.2, for any $p > \frac{1}{3}$, we get

$$\beta(x_H + 1, y_H - 1, z_H) < \beta(x_H, y_H, z_H) < \beta(x_H, y_H - 1, z_H + 1),$$

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. As before, Lemma A.2 implies that for any $p > \frac{1}{3}$,

$$\beta(x_H, y_H + 1, z_H - 1) < \beta(x_H, y_H, z_H) < \beta(x_H - 1, y_H + 1, z_H),$$

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 the continuity of $Eu(g_2) - Eu(i)$ with respect to q_D.

Lemma B.1. Let $\beta : \mathbb{Z}^3 \to [0, 1]$ for given $p > \frac{1}{3}$ be defined as in the main text. Then

$$\beta(x, y, z) = \beta(x + r, y + r, z + r) \quad \forall x, y, z, r \in \mathbb{Z}$$

and $Image(\beta) \subset [0, 1]$ has accumulation points precisely at $0, 1$ and at $\frac{\left(\frac{2p}{1-p}\right)^r}{1+\left(\frac{2p}{1-p}\right)^r}$, for any $r \in \mathbb{Z}$.

Proof. The statement on additive invariance is trivial. Accumulation points at 0 resp. 1 arise from sequences $\beta(n, 0, 0)$ resp. $\beta(0, 0, n)$ with $n \to \infty$. Furthermore, for any $r \in \mathbb{Z}$ the sequences $\beta(s, 0, s + r)$ resp. $\beta(s, 1, s + r)$ are decreasing in s and

$$\lim_{s\to\infty} \beta(s, 0, s + r) = \frac{\left(\frac{2p}{1-p}\right)^r}{1 + \left(\frac{2p}{1-p}\right)^r} = \lim_{s\to\infty} \beta(s, 1, s + r) \quad \forall r \in \mathbb{Z}.$$

To see that these are all accumulation points, take an arbitrary non-stationary but convergent sequence $\beta(x_n, y_n, z_n)$, $n \in \mathbb{N}$. Clearly, it must hold that $x_n + y_n + z_n \to \infty$ due to non-stationarity. By symmetry of β in the last two components, we can without loss of generality assume that $y_n \leq z_n$. If $|z_n - x_n|$ is unbounded, the sequence converges either to 0 or to 1. However, if $|z_n - x_n|$ is bounded, by additive invariance we can assume that $y_n = 0$ for sufficiently large n, and hence for sufficiently large n the sequence is a subsequence of the stylized sequence mentioned above.

Proof of Theorem 2. To prove Part a) of Theorem 2, we rely on the structural insights from Lemma B.1. We again 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:

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

ii) For any jury size n, the threshold profile (x_H, y_H, z_H) satisfies $y_H < 2$.

Note that in case i), q_H is necessarily bounded away from any accumulation point of $Image\,(\beta)$ as for jury size $\tilde{n}$ we have

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

On the other hand, in case ii) we have $q_H \in \left(\beta\left(x_H+1, z_H, z_H\right), \frac{\left(\frac{2p}{1-p}\right)^{z_H - x_H}}{1+\left(\frac{2p}{1-p}\right)^{z_H - x_H}}\right]$. By additive invariance of β and Lemma B.1, it is therefore enough to prove the claim for the following two types of threshold profiles: (x_H, y_H, z_H) with $y_H \geq 2$ and constant $z_H - x_H$, $z_H - y_H$, where x_H, y_H, z_H can be arbitrarily large (case i) and threshold profiles (x_H, y_H, z_H) with $y_H < 2$ and constant $z_H - x_H \in \mathbb{Z}$, where z_H and x_H can be arbitrarily large (case ii).

Fix m and q_H resp. (x_H, y_H, z_H) for some given committee size n. To abbreviate notation, write $\gamma := \frac{2p}{1-p} \geq 2$.

The case $y_H \geq 2$: 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 guarantees the existence of at least m conflict profiles, namely profiles $Piv_3\,(r) + (0,0,1)$ 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.3 for the definition of $Piv_3\,(r)$).

Computing the difference in expected utilities between truthtelling and deviating for each pivotal profile (cf. Lemma A.3) yields

$$
\begin{aligned}
Eu(g_2|Piv_1(r)) - Eu(i|Piv_1(r)) &= -P\left[I|g_2\right]\cdot P\left[\left(x_H+r,y_H-2\left(r+1\right),z_H+r+1\right)|I\right]\cdot q_D \\
&\quad +P\left[G_1|g_2\right]\cdot P\left[\left(x_H+r,y_H-2\left(r+1\right),z_H+r+1\right)|G_1\right]\cdot\left(1-q_D\right) \\
&\quad +P\left[G_2|g_2\right]\cdot P\left[\left(x_H+r,y_H-2\left(r+1\right),z_H+r+1\right)|G_2\right]\cdot\left(1-q_D\right) \\
&= -\frac{1-p}{2}\cdot\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^{n-1}\cdot\gamma^{x_H+r}}{(x_H+r)!\cdot(y_H-2(r+1))!\cdot(z_H+r+1)!}\cdot\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}} \\
&\quad +\frac{1-p}{2}\cdot\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^{n-1}\cdot\gamma^{y_H-2(r+1)}}{(x_H+r)!\cdot(y_H-2(r+1))!\cdot(z_H+r+1)!}\cdot\frac{\gamma^{x_H+\frac{y_H}{2}-m-1}}{\gamma^{x_H+\frac{y_H}{2}-m-1}+\gamma^{2m+1}+\gamma^{z_H+\frac{y_H}{2}-m}} \\
&\quad +p\cdot\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^{n-1}\cdot\gamma^{z_H+r+1}}{(x_H+r)!\cdot(y_H-2(r+1))!\cdot(z_H+r+1)!}\cdot\frac{\gamma^{x_H+\frac{y_H}{2}-N-1}}{\gamma^{x_H+\frac{y_H}{2}-m-1}+\gamma^{2m+1}+\gamma^{z_H+\frac{y_H}{2}-m}} \\
&= \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)} \\
&\quad \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) \\
Eu(g_2|Piv_1^T(r)) - Eu(i|Piv_1^T(r)) &= \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)} \\
&\quad \cdot\left(-\gamma^{r+2m+1}+\gamma^{\frac{3}{2}y_H-2(r+1)-m}\right) \\
Eu(g_2|Piv_2(s)) - Eu(i|Piv_2(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^{n}\cdot\gamma^{x_H}}{(x_H-s)!\cdot(y_H+2s-1)!\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)} \\
&\quad \cdot\left(-\gamma^{-s+2m+1}+\gamma^{\frac{3}{2}y_H+2s-m-2}\right) \\
Eu(g_2|Piv_2^T(s)) - Eu(i|Piv_2^T(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^{n}\cdot\gamma^{x_H}}{(x_H-s)!\cdot(y_H+2s-1)!\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)} \\
&\quad \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) \\
Eu(g_2|Piv_3(r)) - Eu(i|Piv_3(r)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^{n}\cdot\gamma^{x_H}}{(x_H+r)!\cdot(y_H-2r-1)!\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)} \\
&\quad \cdot\left(-\gamma^{r+2m+1}+\gamma^{\frac{3}{2}y_H-2r-m-2}\right) \\
Eu(g_2|Piv_3^T(r)) - Eu(i|Piv_3^T(r)) &= 0 \\
Eu(g_2|Piv_4(s)) - Eu(i|Piv_4(s)) &= \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-1)!\cdot\left(\gamma^{x_H+\frac{y_H}{2}-m-1}+\gamma^{2m+1}+\gamma^{z_H+\frac{y_H}{2}-m}\right)} \\
&\quad \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) \\
Eu(g_2|Piv_4^T(s)) - Eu(i|Piv_4^T(s)) &= 0.
\end{aligned}
$$

The proof now consists of estimating the total sum of utilities from all these profiles by looking at particular combinations of profiles one after another.

Profiles $Piv_2(s)$ incentivize truthtelling given that $y_H \geq 2m+2$.

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

$$\frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{(x_H - s)! \cdot (y_H + 2s)! \cdot (z_H - s)! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)}$$
$$\cdot (z_H + y_H + s) \cdot \left(-\gamma^{-s+2m+1} - \gamma^{z_H + \frac{y_H}{2} - s - m} + \gamma^{\frac{3}{2} y_H + 2s - m - 1} + \gamma^{z_H + \frac{y_H}{2} - s - m - 1}\right).$$

For sufficiently large x_H, y_H, z_H relative to s and relative to the (constant!) differences of each of these, the factor does barely depend on s while the polynomial is increasing in s. So the most negative impact of such profiles is given for the minimal feasible value of s (if any is feasible at all), namely $s = 0$.

To make up for the potential loss from profiles $Piv_2^T(s)$ and $Piv_4(s)$, look at utilities coming from profiles $Piv_1(r)$ for values $r < \left\lfloor \frac{y_H - 1}{2} \right\rfloor - m - 3$. They all incentivize truthtelling, so focus on $r = 0, ..., 2\left\lfloor \frac{\Delta_H + 1}{3} \right\rfloor$ where $2\left\lfloor \frac{\Delta_H + 1}{3} \right\rfloor \ll \left\lfloor \frac{y_H - 1}{2} \right\rfloor - m - 3$ for sufficiently large y_H. The difference in expected utility is given as

$$\frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{(x_H + r)! \cdot (y_H - 2(r+1))! \cdot (z_H + r + 1)! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)}$$
$$\cdot \left(-\gamma^{r+2m+1} - \gamma^{z_H + \frac{y_H}{2} + r - m} + \gamma^{\frac{3}{2} y_H - 2(r+1) - m - 1} + \gamma^{z_H + \frac{y_H}{2} + r - m + 1}\right).$$

Again, for large x_H, y_H, z_H relative to r, the factor barely depends on r while the polynomial is increasing in r. So the least positive impact of such profiles is given for the minimal feasible value of r, namely $r = 0$.

Summing up $Piv_2^T(0)$, $Piv_4(0)$ and $Piv_1(0)$ counted twice (as we consider twice as many profiles of type Piv_1 as of types Piv_2^T and Piv_4 and all Piv_1-type profiles are estimated from below by $Piv_1(0)$) finally yields

$$\frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{x_H! \cdot y_H! \cdot z_H! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)}$$
$$\cdot (z_H + y_H) \cdot \left(-\gamma^{2m+1} - \gamma^{z_H + \frac{y_H}{2} - m} + \gamma^{\frac{3}{2} y_H - m - 1} + \gamma^{z_H + \frac{y_H}{2} - m - 1}\right)$$
$$+ \quad 2 \cdot \frac{(n-1)! \cdot \left(\frac{1-p}{2}\right)^n \cdot \gamma^{x_H}}{x_H! \cdot (y_H - 2)! \cdot (z_H + 1)! \cdot \left(\gamma^{x_H + \frac{y_H}{2} - m - 1} + \gamma^{2m+1} + \gamma^{z_H + \frac{y_H}{2} - m}\right)}$$
$$\cdot \left(-\gamma^{2m+1} - \gamma^{z_H + \frac{y_H}{2} - m} + \gamma^{\frac{3}{2} y_H - 2 - m - 1} + \gamma^{z_H + \frac{y_H}{2} - m + 1}\right).$$

The factor in front of the first term (whose polynomial is negative) is approximately as large as the factor in front of the second term (whose polynomial is positive), given that x_H, y_H, z_H are sufficiently large. To compare the polynomials, for sufficiently large values of y_H, z_H it is enough to compare the terms

containing any of these numbers. Here, as $\gamma \geq 2$, the absolute value of the second polynomial dominates the first one by a factor of at least 2. So the overall sum is positive.

For any $r < \left\lfloor \frac{y_H - 1}{2} \right\rfloor - m - 3$, $Piv_1^T(r)$ incentivizes truthtelling.

Summing up expected utilities of profiles $Piv_1(r-1)$, $Piv_1^T(r-1)$ and $Piv_3(r)$ for $r = \left\lfloor \frac{y_H - 1}{2} \right\rfloor - m - 2, ..., \left\lfloor \frac{y_H - 1}{2} \right\rfloor$ with $y_H \geq 2m + 2$ yields

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

Finally, if y_H is even, summing up expected utilities from profiles $Piv_1(r)$, $Piv_1^T(r)$ with $r = \frac{y_H}{2} - 1$ for $y_H \geq 2m + 2$ yields

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

The case $y_H < 2$: Look at $q_D = \beta\left(x_H - \left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor, y_H + 2\left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor - 4, z_H - \left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor\right) = \frac{\gamma^{y_H + 3\left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor - 4} + \gamma^{z_H}}{\gamma^{x_H} + \gamma^{y_H + 3\left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor - 4} + \gamma^{z_H}}$ with y_H being either 0 or 1. This choice guarantees conflict for any profile of type $Piv_4(s) + (0, 0, 1)$ with $s \leq \left\lfloor \frac{z_H - y_H - 1}{3} \right\rfloor - 2$, so choosing the committee size sufficiently large guarantees m conflict profiles as z_H gets arbitrarily large while $y_H < 2$. Note that no Pivotal profiles of type Piv_1, Piv_1^T or Piv_3, Piv_3^T exist (except possibly $Piv_3(0) = Piv_2(0)$, $Piv_3^T(0) = Piv_2^T(0)$).

Computing the difference in expected utilities between truthtelling and deviating for each type of pivotal profiles yields

$$
\begin{aligned}
Eu(g_2|Piv_2(s)) - Eu(i|Piv_2(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_H}}{(x_H-s)!\cdot(y_H+2s-1)!\cdot(z_H-s)!\cdot\left(\gamma^{x_H}+\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}+\gamma^{z_H}\right)}\\
&\quad\cdot\left(-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-s-4}-\gamma^{z_H-s}+\gamma^{y_H+2s-1}+\gamma^{z_H-s+1}\right)\\
Eu(g_2|Piv_2^T(s)) - Eu(i|Piv_2^T(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_H}}{(x_H-s)!\cdot(y_H+2s-1)!\cdot(z_H-s)!\cdot\left(\gamma^{x_H}+\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}+\gamma^{z_H}\right)}\\
&\quad\cdot\left(-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-s-4}+\gamma^{y_H+2s}\right)\\
Eu(g_2|Piv_4(s)) - Eu(i|Piv_4(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_H}}{(x_H-s)!\cdot(y_H+2s)!\cdot(z_H-s-1)!\cdot\left(\gamma^{x_H}+\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}+\gamma^{z_H}\right)}\\
&\quad\cdot\left(-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-s-4}+\gamma^{y_H+2s}\right)\\
Eu(g_2|Piv_4^T(s)) - Eu(i|Piv_4^T(s)) &= 0.
\end{aligned}
$$

Summing up $Piv_2(0), Piv_2^T(0)$ (if they exist) yields

$$
\begin{aligned}
&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_H}}{x_H!\cdot(y_H-1)!\cdot z_H!\cdot\left(\gamma^{x_H}+\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}+\gamma^{z_H}\right)}\\
&\cdot\left(-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}-\gamma^{z_H}+\gamma^{y_H-1}+\gamma^{z_H+1}-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}+\gamma^{y_H}\right)\\
>\quad & 0.
\end{aligned}
$$

Summing up $Piv_2(s), Piv_2^T(s), Piv_4(s-1)$ for $1 \le s \le \left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor$ yields

$$
\begin{aligned}
&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_H}}{(x_H-s)!\cdot(y_H+2s-1)!\cdot(z_H-s)!\cdot\left(\gamma^{x_H}+\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}+\gamma^{z_H}\right)}\\
&\cdot\left(\left(-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-s-4}-\gamma^{z_H-s}+\gamma^{y_H+2s-1}+\gamma^{z_H-s+1}-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-s-4}+\gamma^{y_H+2s}\right)\right.\\
&\left.+\ \frac{y_H+2s-1}{x_H-s+1}\left(-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-s-3}+\gamma^{y_H+2s-2}\right)\right).
\end{aligned}
$$

By choosing committee size sufficiently large, the quotient in front of the second polynomial is only marginally larger than 1 (if at all). Taking this quotient as being equal to 1, the remaining polynomial is larger or equal than

$$
\gamma^{-s}\cdot\left(-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}-\gamma^{z_H}+\gamma^{z_H+1}-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-4}-\gamma^{y_H+3\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor-3}\right)>0.
$$

Finally, profiles $Piv_4\left(\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor\right), Piv_4^T\left(\left\lfloor\frac{z_H-y_H-1}{3}\right\rfloor\right)$ (weakly) incentivize truthtelling.

This completes the proof of Part a).

Part b) follows from Part a) in the following way: Note that, for fixed m, in Part a) we found an open interval $B_\delta(q_H)$ around any value of $q_H \in (0,1)$ such that for any $\tilde{q}_H \in B_\delta(q_H)$ there exist at least m conflict profiles between $\tilde{q}_H$ and $\hat{q}_D(\tilde{q}_H)$, assuming that the committee has at least $\hat{n}(q_H)$ members. This is clear for the case where $y_H \geq 2$ by Lemma B.1. For the case $y_H < 2$ this follows from the fact that q_D was chosen independent of m and, up to 3-periodicity, independent of n. Part b) then follows from compactness of $[\epsilon, 1-\epsilon]$ for any $\epsilon > 0$.

Theorem B.1. *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 . The proof of Theorem B.1 is virtually identical to the proof of Theorem 1 and therefore omitted.

Theorem B.2. *Let doves have critical mass.*

a) 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.

b) Fix some $m \in \mathbb{N}$ and some $\epsilon > 0$. Then there exists some threshold committee size $\hat{\hat{n}}$ s.t. for any committee size $n \geq \hat{\hat{n}}$, any $q_D \in [\epsilon, 1-\epsilon]$ 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 . 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 dictinction 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 $q_H = \beta(x_D + m, y_D - 2m, z_D + m) = \frac{\gamma^{y_D-2m} + \gamma^{z_D+m}}{\gamma^{x_D+m} + \gamma^{y_D-2m} + \gamma^{z_D+m}}$, assuming $y_D \geq 2m$.

Computing the difference in expected utilities between truthtelling and deviating for each type of pivotal profiles yields

$$
\begin{aligned}
Eu(i|Piv_1(r)) - Eu(g_2|Piv_1(r)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D+r)!\cdot(y_D-2(r+1))!\cdot(z_D+r+1)!\cdot\left(\gamma^{x_D+m}+\gamma^{y_D-2m}+\gamma^{z_D+m}\right)}\\
&\quad\cdot\left(+\gamma^{y_D+r-2m+1}-\gamma^{y_D-2(r+1)+m}\right)\\
Eu(i|Piv_1^T(r)) - Eu(g_2|Piv_1^T(r)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D+r)!\cdot(y_D-2(r+1))!\cdot(z_D+r+1)!\cdot\left(\gamma^{x_D+m}+\gamma^{y_D-2m}+\gamma^{z_D+m}\right)}\\
&\quad\cdot\left(+\gamma^{y_D+r-2m+1}-\gamma^{y_D-2(r+1)+m}\right)\\
Eu(i|Piv_2(s)) - Eu(g_2|Piv_2(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s-1)!\cdot(z_D-s)!\cdot\left(\gamma^{x_D+m}+\gamma^{y_D-2m}+\gamma^{z_D+m}\right)}\\
&\quad\cdot\left(+\gamma^{y_D-s-2m+1}+\gamma^{z_D-s+m+1}-\gamma^{y_D+2s+m-1}-\gamma^{z_D-s+m}\right)\\
Eu(g_2|Piv_2^T(s)) - Eu(i|Piv_2^T(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s-1)!\cdot(z_D-s)!\cdot\left(\gamma^{x_D+m}+\gamma^{y_D-2m}+\gamma^{z_D+m}\right)}\\
&\quad\cdot\left(+\gamma^{y_D-s-2m+1}+\gamma^{z_D-s+m+1}-\gamma^{y_D+2s+m-1}-\gamma^{z_D-s+m}\right)\\
Eu(i|Piv_3(r)) - Eu(g_2|Piv_3(r)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D+r)!\cdot(y_D-2r-1)!\cdot(z_D+r)!\cdot\left(\gamma^{x_D+m}+\gamma^{y_D-2m}+\gamma^{z_D+m}\right)}\\
&\quad\cdot\left(+\gamma^{y_D+r-2m+1}+\gamma^{z_D+r+m+1}-\gamma^{y_D-2r+m-1}-\gamma^{z_D+r+m}\right)\\
Eu(i|Piv_3^T(r)) - Eu(g_2|Piv_3^T(r)) &= 0\\
Eu(i|Piv_4(s)) - Eu(g_2|Piv_4(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s)!\cdot(z_D-s-1)!\cdot\left(\gamma^{x_D+m}+\gamma^{y_D-2m}+\gamma^{z_D+m}\right)}\\
&\quad\cdot\left(+\gamma^{y_D-s-2m+1}+\gamma^{z_D-s+m+1}-\gamma^{y_D+2s+m}-\gamma^{z_D-s+m-1}\right)\\
Eu(i|Piv_4^T(s)) - Eu(g_2|Piv_4^T(s)) &= 0.
\end{aligned}
$$

Summing up expected utilities of profiles $Piv_1(r), Piv_1^T(r), Piv_3(r)$ for $r \leq m-1$ yields

$$
\begin{aligned}
&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D+r)!\cdot(y_D-2r-1)!\cdot(z_D+r+1)!\cdot\left(\gamma^{x_D+m}+\gamma^{y_D-2m}+\gamma^{z_D+m}\right)}\\
&\left(2(y_D-2r-1)\left(+\gamma^{y_D+r-2m+1}-\gamma^{y_D-2(r+1)+m}\right)\right.\\
&\left.+(z_D+r+1)\left(+\gamma^{y_D+r-2m+1}+\gamma^{z_D+r+m+1}-\gamma^{y_D-2r+m-1}-\gamma^{z_D+r+m}\right)\right).
\end{aligned}
$$

This expression is positive for all $r \leq m-1$. All remaining profiles incentivize truthtelling.

The case $y_D < 2$: Fix $q_H = \beta(x_D+1, y_D-1, z_D) = \frac{\gamma^{y_D-1}+\gamma^{z_D}}{\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}}$ with y_D being either 0 or 1. This choice guarantees conflict for any profile of type $Piv_2(s)+(1,0,0)$ with $s \leq \left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor - 1$, so choosing the committee size sufficiently large guarantees m conflict profiles. Note that no pivotal profiles of type Piv_1, Piv_1^T or Piv_3, Piv_3^T exist (except possibly $Piv_3(0) = Piv_2(0)$, $Piv_3^T(0) = Piv_2^T(0)$).

Computing the difference in expected utilities between truthtelling and deviating for each type of pivotal profiles yields

$$
\begin{aligned}
Eu(i|Piv_2(s)) - Eu(g_2|Piv_2(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s-1)!\cdot(z_D-s)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\quad\cdot\left(+\gamma^{y_D-s}-\gamma^{y_D+2s}\right)\\
Eu(i|Piv_2^T(s)) - Eu(g_2|Piv_2^T(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s-1)!\cdot(z_D-s)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\quad\cdot\left(+\gamma^{y_D-s}-\gamma^{y_D+2s}\right)\\
Eu(i|Piv_4(s)) - Eu(g_2|Piv_4(s)) &= \frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s)!\cdot(z_D-s-1)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\quad\cdot\left(+\gamma^{y_D-s}+\gamma^{z_D-s+1}-\gamma^{y_D+2s+1}-\gamma^{z_D-s}\right)\\
Eu(i|Piv_4^T(s)) - Eu(g_2|Piv_4^T(s)) &= 0.
\end{aligned}
$$

Summing up $Piv_4(s), Piv_2(s), Piv_2^T(s)$ for $s \leq \left\lfloor \frac{z_D-y_D+1}{3} \right\rfloor - 4$ yields

$$
\begin{aligned}
&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s)!\cdot(z_D-s-1)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\cdot\left(+\gamma^{y_D-s}+\gamma^{z_D-s+1}-\gamma^{y_D+2s+1}-\gamma^{z_D-s}\right)\\
&+2\cdot\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s-1)!\cdot(z_D-s)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\cdot\left(+\gamma^{y_D-s}-\gamma^{y_D+2s}\right)\\
=\ &\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{(x_D-s)!\cdot(y_D+2s)!\cdot(z_D-s)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\cdot\Big((z_D-s)\left(+\gamma^{y_D-s}+\gamma^{z_D-s+1}-\gamma^{y_D+2s+1}-\gamma^{z_D-s}\right)\\
&+2\,(y_D+2s)\left(+\gamma^{y_D-s}-\gamma^{y_D+2s}\right)\Big)\\
\geq\ &\gamma^{y_D-s}+\gamma^{z_D-s+1}-\gamma^{y_D+2s+1}-\gamma^{z_D-s}+\gamma^{y_D-s}-\gamma^{y_D+2s}\\
\geq\ &\gamma^{\frac{2}{3}z_D+\frac{1}{3}y_D-\frac{1}{3}}\cdot\left(+\gamma^5-\gamma^{-6}-\gamma^4-\gamma^{-7}\right)\\
\geq\ &0.
\end{aligned}
$$

Next, profiles $Piv_4\left(\left\lfloor \frac{z_D-y_D+1}{3} \right\rfloor - 3\right)$ and $Piv_2\left(\left\lfloor \frac{z_D-y_D+1}{3} \right\rfloor - t\right), Piv_2^T\left(\left\lfloor \frac{z_D-y_D+1}{3} \right\rfloor - t\right)$ for $t = 0, \ldots, 3$ add up to

$$
\begin{aligned}
&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{\left(x_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+3\right)!\cdot\left(y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-6\right)!\cdot\left(z_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+2\right)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\left(+\gamma^{y_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+3}+\gamma^{z_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+4}-\gamma^{y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-5}-\gamma^{z_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+3}\right)\\
+\quad 2\cdot&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{\left(x_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor\right)!\cdot\left(y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-1\right)!\cdot\left(z_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor\right)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\left(+\gamma^{y_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor}-\gamma^{y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor}\right)\\
+\quad 2\cdot&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{\left(x_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+1\right)!\cdot\left(y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-3\right)!\cdot\left(z_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+1\right)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\left(+\gamma^{y_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+1}-\gamma^{y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-2}\right)\\
+\quad 2\cdot&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{\left(x_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+2\right)!\cdot\left(y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-5\right)!\cdot\left(z_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+2\right)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\left(+\gamma^{y_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+2}-\gamma^{y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-4}\right)\\
+\quad 2\cdot&\frac{(n-1)!\cdot\left(\frac{1-p}{2}\right)^n\cdot\gamma^{x_D}}{\left(x_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+3\right)!\cdot\left(y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-7\right)!\cdot\left(z_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+3\right)!\cdot\left(\gamma^{x_D+1}+\gamma^{y_D-1}+\gamma^{z_D}\right)}\\
&\left(+\gamma^{y_D-\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor+3}-\gamma^{y_D+2\left\lfloor\frac{z_D-y_D+1}{3}\right\rfloor-6}\right).
\end{aligned}
$$

Up to the additional factor 2, all the coefficients are virtually identical for sufficiently large committees, so it suffices to compare the polynomials which are larger or equal than

$$\gamma^{\frac{2}{3}z_D+\frac{1}{3}y_D-\frac{1}{3}}\left(\gamma^4-\gamma^3-2\gamma^1-2\gamma^{-1}-2\gamma^{-3}-2\gamma^{-4}-2\gamma^{-5}\right)>0.$$

Finally, all remaining profiles of type $Piv_4(s)$ incentivize truthtelling. This proves Part a).

Part b) follows from Part a) in the same way as for Theorem 2.

References

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

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

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

[4] J. Elster, *The Market and the Forum: Three Varieties of Political Theory, In Deliberative Democracy: Essays on Reason and Politics.* Cambridge, MIT Press, 1997.

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

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

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

[8] J. Habermas, *Moral Consciousness and Communicative Action.* Cambridge, MIT Press, 1990.

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

[10] M. Le Quement, "Communication compatible voting rules." mimeo, University of Bonn, 2012.

[11] M. Le Quement and V. Yokeeswaran, "Deliberation, conflict and unanimity." mimeo, University of Bonn, 2011.

[12] B. Manin, "On legitimacy and political deliberation," *Political Theory*, vol. 15, pp. 338–368, 1987.

[13] 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.

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

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