THE IMPORTANCE OF EXPERTISE IN GROUP DECISIONS

ALEXANDER LUNDBERG

Abstract. Prior to a collective binary choice, members of a group receive binary signals correlated with the better option. Expanding membership may provide no benefit, but expertise is everywhere beneficial. If the group ignores any statistical dependence among the signals, as through majority vote, an expert may perform better than the group. If the group accounts for dependence, a relatively expert member puts an upper bound on the probability of a false belief. The bound holds for any group size and signal distribution. Furthermore, a population investing in expertise is better off cultivating a small mass of elites than adopting an egalitarian policy.

Keywords: Bayesian Learning, Expert Advice, Condorcet Jury Theorem, Committee Voting. *JEL*: D71, D72, D83.

I. Introduction

Expert advice pervades modern society. Patients visit doctors for diagnoses. Journalists cite academics in their stories. Firms hire consultants to evaluate mergers. At its core, expertise is information, and we hope experts make more informed decisions.[1] Furthermore, we hope more informed decisions improve welfare. Economic studies offer encouragement. For example, Bloom et al. (2013) find expert management practices raise firm profits, and Bronnenberg et al. (2015) find informed shoppers are more likely to buy less expensive store brand versions of largely equivalent products.

If expertise is good, and more is better, then groups of the most distinguished experts presumably make the best decisions. However, the Marquis de Condorcet famously discounted the necessity of expertise in a remarkable insight over two centuries ago. If enough marginally informed people decide an outcome by majority vote, the group will collectively arrive at the right decision (Condorcet 1785). Condorcet's proposal has broad implications for society. It can motivate a diverse range of institutions, from democracy itself, to an expressive function of law (Dharmapala and McAdams 2003), to hierarchical decision making in firms (Katzner 1995). Because of its wide applicability, the Condorcet Jury Theorem (CJT) continues to command interdisciplinary research attention. Economists, political scientists, and mathematicians in particular have devoted several decades to its delineation.

In its classical version, the CJT presents a group facing a binary decision. Each member has a signal correlated with the better of the two options. As the group size grows, a majority vote identifies the better option with probability nearing one. The result hinges on the assumption of statistical independence. No signal provides information on any other. In practice, however, individuals can share common information, culture, ideology, etc. Such

Date: September 6, 2017.

[1]Skill refinement is another form of expertise, but decisional ability is the focus here.

commonalities weaken the case for independence, but the CJT is robust to a variety of extensions. For example, as long as the average correlation between signals goes to zero as group size grows, the result holds (Ladha 1992). Furthermore, the result holds for a broad class of dependent signals (Peleg and Zamir 2012), and negative correlation can even be helpful (Berg 1993).

Under a variety of conditions, then, a large enough group will continue to make the right decision with almost perfect accuracy. Is expertise therefore overrated? This article provides a "not so fast" answer. Under arbitrary dependence, an infinitely sized group may fail to perform any better than a single member. Importantly, the failure can occur whether the group effectively ignores the dependence via majority vote, as in most of the literature, or members pool their information to update as a collectively rational agent. A majority vote tends to be the more dangerous approach though. If the group ignores dependence, expanding membership can even be harmful. A sufficiently large number of relatively uninformed members with highly correlated votes can always outweigh a more informed contingent. Since one way to eliminate such a risk is to restrict group membership, a good rule of thumb for a risk-averse (or ambiguity-averse, more accurately) authority is to delegate key decisions to experts.

If a group does account for dependence, expertise has a clearly defined benefit. A single expert places an upper bound on the probability of a mistaken belief. The bound is attractive because it applies under any group size and any signal distribution.

The results are motivated through a graphical interpretation of statistical dependence. To get a sense for the classical setting, consider what the assumption of independence means for a sequence of signals. Let $\mathcal{A}$ and $\mathcal{B}$ denote the two states of the world. Each group member i receives a signal, a_i or b_i. The signals are informative, symmetric across members, and symmetric around the state of the world, meaning $P(a_i|\mathcal{A}) = P(b_i|\mathcal{B}) = p > 1/2 \ \forall i \in \mathbb{N}^+ \equiv \{1, 2, 3, ...\}$. Figure 1 provides a graphical depiction of independence conditional on state $\mathcal{A}$. Without loss of generality, the events a_1 and b_1 are grouped into contiguous blocks.

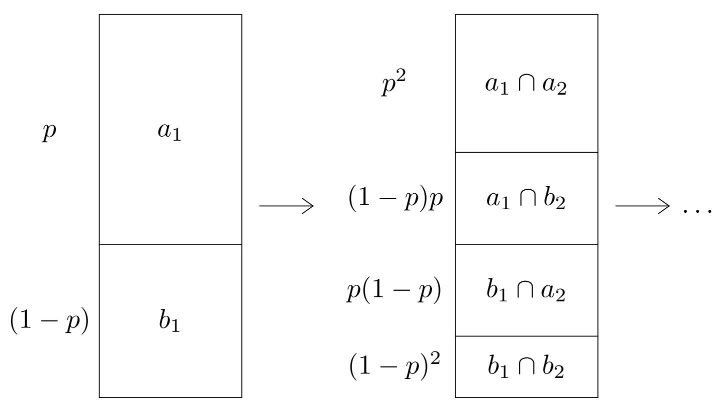

FIGURE 1. Independence Conditional on $\mathcal{A}$.

The picture reveals the precision required by the assumption. Informally speaking, the only permissible sequence requires the signal for each group member to subdivide the previous signals into an event of area p for the correct inference and area $(1-p)$ for the incorrect one. Any other of the infinite possible sequences would violate independence.

Extensions of the CJT allowing for dependence generally require a type of symmetry between group members (Peleg and Zamir 2012). The graphical approach is useful because it yields conclusions under any type of dependence, and that flexibility is useful because a number of social processes can produce different correlations between signals. For one, network structure can shape learning and information diffusion (Jackson et al. 2017). Group members might also converse with each other before voting, as in DeGroot (1974). Allowing for arbitrary dependence is a way to capture such learning prior to a decision. Of course, proper inference does require a knowledge of the underlying dependence structure. A majority vote effectively ignores any dependence, which can distort beliefs. Again, one advantage of choosing an expert over a group is the reduction of belief distortion caused by correlation neglect.[2]

If group size provides no guarantee of a good decision, but expertise does, then democratic participation might be counterproductive. Meetings and elections can impose hefty transaction costs, and elitism is thereby a less expensive policy than democratic engagement. Furthermore, consider the problem of social investment in human capital. Even in the original, highly democratic setting envisioned by Condorcet, a strategy of elitism is optimal. Though feasible in technicality, a policy of universal education is only superior when parameter values are grossly implausible; the benefit of a correct decision must be many orders of magnitude higher than investment and participation costs for even small population sizes. Rather than invest in universal expertise, society does better by cultivating a few elites and relying on their wisdom.

While the results mostly dampen enthusiasm for organizational democracies, certain types of statistical dependence offer a point of optimism. For example, three individuals, each of whom is more likely to be wrong than right, can perfectly determine the state of the world if they share just the right dependence among their signals (but they must understand the nature of the correlation to draw the right inference). If society could identify such dependencies, or create them by investment, good decisions might be comparatively cheap and frequent. Additionally, expertise is never cheap, often requiring years of continuous work to develop, nor is it a panacea—even famous experts can and do fail spectacularly in their predictions. Still, the analysis provides a sobering overall picture of democratic participation.

The next section briefly describes the related literature. Section III presents the basic model, and Section IV presents the main results on the value of expertise. Section V discusses the voter participation and investment decisions. Finally, Section VI presents an optimistic result, which demands no expertise, while Section VII concludes. All proofs are left to the Appendix.

[2] A number of psychological pitfalls can also beset group decision making (e.g., Bénabou [2013], Sunstein [2005]). This article emphasizes the importance of expertise in a rational setting, but groupthink can devalue its function in practice. On the other hand, Charness and Sutter (2012) cite a body of experimental evidence in which groups are less susceptible to cognitive biases than individuals in games of strategy.

II. Related Literature

The literature on the CJT is vast, and this section is not a complete review. Rather, it explains how this article fits into the current literature.

In its most basic form, the CJT assumes each of n individuals has a signal with probability $p > 1/2$ of indicating the better of two outcomes.[3] The signals are independent, and the outcome is decided by majority vote. Most formulations of the CJT assume n is odd. If tie votes are decided by a coin flip, an even group size makes no material difference to any conclusion, but combining even and odd results introduces unwieldy non-monotonicity. For convenience, this article largely follows the convention of an odd group size. Define

$$F(p,n) \equiv \sum_{k=\frac{(n+1)}{2}}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

as the probability of a correct majority vote for accuracy p and group size n. Then,

Proposition 1. If $p > 1/2$ and signals are mutually independent, $\lim_{n\to\infty} F(p,n) = 1$. Further, if n is odd, the sequence $\{F(p,n)\}_{n=1}^{\infty}$ is monotonic.

For a proof of *Proposition 1*, see Boland (1989).

Many studies aim to relax the assumptions of the classical CJT in three principal ways.

II.1. **Homogeneity.**

Not every group member needs an identical p. Ben-Yashar and Paroush (2000) provide a finite CJT in which a majority vote performs better than a vote from a randomly selected group member, assuming each signal accuracy is greater than one half. The signals do need to be bounded above one half for the CJT to hold (Paroush 1998), but adding one informed member (with $p_i > 1/2$) and one uninformed member ($p_i = 1/2$) always improves the the performance of the group (Ben-Yashar and Zahavi 2011). Surprisingly, the CJT can fail when members are more likely to be incorrect than correct in one of two states of nature, even if they are correct on average over both states (Ben-Yashar 2014). Most work focuses on sufficient conditions, but Berend and Paroush (1998) provide a necessary and sufficient condition for the asymptotic CJT under heterogenous signals. Lindner (2008) further addresses the case of unequal voting weights.

II.2. **Independence.**

Nitzan and Paroush (1984) were the first to address the importance of independence, showing simple majority is not always the best voting rule without it. Boland (1989) and Berg (1993, 1994) prove the asymptotic CJT for cases of dependent voters with a leader who influences the others. Ladha (1995) puts forward several correlated distributions for which the theorem still holds. In the examples, positive correlation reduces competency but negative correlation improves it. For deriving probability distributions in applied work, Kaniovski (2010) offers conditions for correlation to be helpful in different solution methods.

[3] While Condorcet first proposed his vision for collective wisdom in 1785, Laplace offered the first mathematical proof of the hypothesis in 1812 (Ben-Yashar and Paroush 2000).

If the average correlation between signals goes to zero as group size grows, the CJT still holds (Ladha 1992). The non-asymptotic CJT also holds for all exchangeable signals—that is, when jurors can be treated symmetrically or "anonymously" (Ladha 1993). Berend and Sapir (2007) provide several equivalent conditions for the non-asymptotic verison of the CJT, while Peleg and Zamir (2012) characterize the asymptotic counterpart under exchangeability.

Levy and Razin (2015) present a different model of beneficial correlation. Voters exist along a continuum with varying strength of policy preference. However, all voters prefer the policy corresponding to the underlying state of the world, which is unknown. Voters receive two, possibly correlated, binary signals indicating the state of the world. Depending on the distribution of preferences, correlation neglect can yield better decisions. Glaeser and Sunstein (2009) also model correlation neglect but over a normally distributed outcome. Group members share their information prior to the decision. When they all ignore correlation, a single member can make a better decision than the group.

II.3. **Strategy.** Sincere voting is an implicit but important assumption of the CJT. Sincerity means each member votes according to her own signal, without regard for strategic interaction. Austen-Smith and Banks (1996) show sincere voting is not always a strategic equilbrium. Wit (1998) shows information aggregation might in fact be more efficient in focal equilibria (see also Myerson 1998). Apparently, unanimity is a uniquely bad voting rule in the strategic context (Feddersen and Pesendorfer 1998), but Coughlan (2000) defends unanimity via two extensions of the model. Meirowitz (2002) shows unanimity can even generate a CJT when voters receive signals from a continuum.

A number of other studies explore voting rules and game theory, but this article will assume voting is sincere. McLennan (1998) provides a justification for the omission of strategic interaction. The properties of any common interest game imply that whenever sincere voting establishes the CJT, there exist symmetric Nash equilibria doing so as well.[4] Additionally, when the optimal voting rule is adopted, sincere voting is always a Nash equilibrium (Ben-Yashar 2006). Laslier and Weibull (2011) also introduce a randomized voting rule which preserves the incentive to vote sincerely.

In addressing the value of expertise, this article straddles the line between relaxing the homogeneity and independence assumptions. Its main contribution is in providing results applicable under arbitrary dependency. It also introduces the questions of participation and investment in the classical setting. The results complement the work of McMurray (2013), who develops a model of costless voting in which participation is incomplete. In the model, an expert voter with superior information has a higher incentive to vote, but her information discourages others from voting, as they prefer to leave the decision to experts.

Other studies are more difficult to categorize but examine various extensions of the CJT. The definition of a correct decision carries another implicit assumption, this time concerning

[4]On a related question, Ahn and Oliveros (2014) show the asymptotic lack of an informational advantage for either joint or separate trials when multiple outcomes are decided. If a sequence of equilibria exists for which the optimal outcome is chosen in the limit for one trial format, such a sequence exists for the other format too.

preferences. Conditional on the state of the world, all individuals prefer the same outcome. Miller (1986) shows the CJT still applies if the correct decision is defined as the outcome preferred by the majority of voters. Several studies endogenize the quality of signals through effort (e.g., Koriyama and Szentes 2009, McCannon and Walker 2016, Mukhopadhaya 2003, Persico 2004). Committee members underprovide information as a public good, which mutes the benefit of information aggregation. For a different approach, Sah and Stiglitz (1998) explore whether a committee vote or a hierarchical system is a better way to make decisions.

A number of theoretical studies examine how to maximize the informational content of expert advice under private biases. For example, Hilger (2016) analyzes the incentives of expert firms to conceal their costs of production. Krishna and Morgan (2001) extend the Crawford and Sobel (1982) model to a setting of two biased experts. Sequentially consulting each one for advice can be informative when they have opposing biases, but the approach is always a bad idea when they have the same bias. Gerardi and Yariv (2008) largely support the result in a different model of costly expertise acquisition. As a final example, Li and Suen (2004) model the inherent tradeoff in delegating a decision to a privately interested expert with a partisan viewpoint. If the prinicipal and expert share the same (opposite) partisanship, they align on more (fewer) decisions, but the principal receives less (more) informative advice. The current article looks at group decisions under a common objective, so the incentive to withhold or distort information is absent.

Finally, Congleton (2007) runs Monte Carlo simulations of a slightly informed electorate and compares their median estimate of a continuous quality measure to an estimate of an expert with a more sophisticated regression technique. In many cases, the electorate performs better than the expert (cf. Surowiecki 2005). This article differs in its theoretical focus on a binary decision.

III. The Model

A group is comprised of n members who receive binary signals correlated with an unknown state of the world, $\mathcal{S} \in \{\mathcal{A}, \mathcal{B}\}$.

III.1. **The Signals.**

Each signal, $s_i \in \{a_i, b_i\}$, is informative but noisy, satisfying

$$\begin{aligned} &1 > P(a_i|\mathcal{A}) = 1 - P(b_i|\mathcal{A}) > 1/2, \\ &1 > P(b_i|\mathcal{B}) = 1 - P(a_i|\mathcal{B}) > 1/2, \text{ and} \\ &a_i \cap a_j \neq a_i \text{ for } i \neq j. \end{aligned} \tag{A1}$$

The last condition requires the signals to be different from each other.

Let $\mathbf{n} \equiv \{1, 2, ..., n\}$ index the group members, and let $\mathbf{k} = \{k_1, k_2, ..., k_{|\mathbf{k}|}\} \subseteq \mathbf{n}$ denote an ordered subset with cardinality $|\mathbf{k}|$, which may be infinite if $n \to \infty$. Define $\mathbb{K}$ as the set of all possible indices for which $\mathbf{k} \neq \emptyset$. Let $S_\mathbf{k}$ denote the set of all possible realizations of the signals indexed by $\mathbf{k}$, and define $\mathbf{s_k} = (s_{k_1}, s_{k_2}, ...)$ as the vector of realized signals, with $\bar{\mathbf{s}}_\mathbf{k}$ denoting its complement. That is, if $\mathbf{s_k} = (a_{k_1}, b_{k_2}, ...)$, then $\bar{\mathbf{s}}_\mathbf{k} = (b_{k_1}, a_{k_2}, ...)$. Also, let $\bar{\mathcal{S}}$ denote the complement of $\mathcal{S} \in \{\mathcal{A}, \mathcal{B}\}$.

Some results will refer to vectors of unrealized signals. Let v_i stand for the unrealized signal i, with $\mathbf{v_k}$ denoting the vector of unrealized signals for a given set $S_\mathbf{k}$.

The traditional Condorcet model assumes symmetry around the true state, implying $P(a_i|\mathcal{A}) = P(b_i|\mathcal{B})\ \forall i \in \mathbb{N}^+$. Accounting for statistical dependence, symmetry is equivalent to

$$P(\mathbf{s_k}|\mathcal{A}) = P(\bar{\mathbf{s}}_\mathbf{k}|\mathcal{B})\ \ \forall \mathbf{k} \in \mathbb{K},\ \mathbf{s_k} \in S_\mathbf{k}. \tag{A2}$$

Assumption (A2) implies the standard equality of $P(a_i|\mathcal{A}) = P(b_i|\mathcal{B})\ \forall i \in \mathbb{N}^+$.

III.2. **The Updates.** The group updates beliefs in one of two ways. In the first scenario, members recognize the correlation between the signals and update as a rational Bayesian:

$$\pi_1 = \frac{P(\mathcal{A}|\cap_{i=1}^n s_i)}{P(\mathcal{B}|\cap_{i=1}^n s_i)} = \frac{P(\cap_{i=1}^n s_i|\mathcal{A})}{P(\cap_{i=1}^n s_i|\mathcal{B})}\frac{P(\mathcal{A})}{P(\mathcal{B})}, \tag{S1}$$

In the second scenario, the members treat the signals as if they were conditionally independent:

$$\pi_2 = \frac{P(\mathcal{A}|\cap_{i=1}^n s_i)}{P(\mathcal{B}|\cap_{i=1}^n s_i)} = \Big[\prod_{i=1}^n \frac{P(s_i|\mathcal{A})}{P(s_i|\mathcal{B})}\Big]\frac{P(\mathcal{A})}{P(\mathcal{B})}, \tag{S2}$$

The "conditional" modifier is dropped in the remaining discussion.

For simplicity, assume the members start with uninformative priors. The priors play no important role in any result, but their equality does abstract from heterogenous initial beliefs.

Lastly, let $\mathcal{G}_n$ denote the set of all possible distributions over the n signals, and let c, e, and f denote the events of placing odds strictly greater than, equal to, or strictly less than one on the true state of the world after. As a mnemonic device, c, e, and f stand for "correct," "equal odds," and "false." Note that correct and false beliefs are not complements if e has non-zero measure.

IV. The Value of Expertise

According to the CJT, group size can compensate for a lack of expertise, but does the claim still hold under statistical dependence? The answer is unfortunately no. In fact, in the limiting case, expertise is essential to an informed decision.

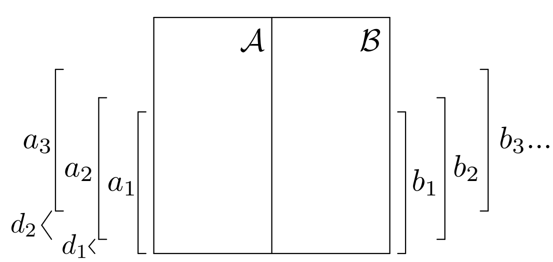

Figure 2. Depiction of Sea of Noise

IV.1. **Sea of Noise.** Although a complete description of every $g \in \mathcal{G}_n$ is not feasible for large n, consider the class depicted by Figure 2, in which each event $a_i|\mathcal{A}$ is obtained by shifting upward the previous $a_{i-1}|\mathcal{A}$ by some non-zero d_{i-1}. Though not pictured, recall

$P(b_i|\mathcal{A}) = 1 - P(a_i|\mathcal{A})$. Without loss of generality, the events can be ordered to satisfy the following assumption for any index $\mathbf{k} \in \mathbb{K}$ and $m \in \mathbb{N}^+$:

$$\begin{aligned}
&P(a_{k_{|\mathbf{k}|}+m} \cap (\bigcup_{i\in\mathbf{k}} a_i)|\mathcal{A}) = p - \sum_{j=k_{|\mathbf{k}|}}^{k_{|\mathbf{k}|}+m-1} d_j, \\
&P(b_{k_{|\mathbf{k}|}+m} \cap (\bigcap_{i\in\mathbf{k}} a_i)|\mathcal{A}) = \sum_{j=k_{|\mathbf{k}|}}^{k_{|\mathbf{k}|}+m-1} d_j, \\
&\sum_{i=1}^{\infty} d_i \equiv \gamma \leq 1 - p.
\end{aligned} \tag{A3}$$

Note how any infinite subset of the sequence retains the same basic structure as the complete set. This convenient feature simplifies various proofs and welfare results.

Lemma 1. $\forall \mathbf{k} \in \mathbb{K}$, $i \in \mathbf{k}$, $\exists\, \delta_i, \sigma_i \in \{\alpha_i, \beta_i\}$—analogs to $d_i, s_i \in \{a_i, b_i\}$—satisfying (A3) $\forall \mathbf{k}' \subset \mathbf{k}$.

The following lemma is also useful.

Lemma 2. $P(\mathbf{s_k}|\mathcal{S}) = P(\mathbf{s_k}|\bar{\mathcal{S}})\ \forall\, \mathbf{s_k} \in S_{\mathbf{k}}$ s.t. $\exists\, i, j \in \mathbf{k}$ for which $s_i \neq s_j$.

Surprisingly, the only informative case occurs when every group member has an identical signal. In every other case, posterior beliefs are equal to priors. More signals therefore only introduce uncertainty.

Suppose, first, the group knows the dependence structure and updates accordingly. Increasing group membership is not entirely bad since the additional signals provide insurance against a false belief. *Proposition 2* summarizes the main conclusions.

Proposition 2. $\forall \mathbf{k} \in \mathbb{K}$, (S1) and (A1)-(A3) $\Rightarrow$

(1) $P(c|\mathbf{v_k}) \leq P(c|v_{\mathbf{k}\setminus\mathbf{y}})$ and $P(f|\mathbf{v_k}) \leq P(f|v_{\mathbf{k}\setminus\mathbf{y}})\ \forall \mathbf{y} \subseteq \mathbf{k}$, and

(2) $\lim_{|\mathbf{k}|\to\infty} P(c|\mathbf{v_k}) = p - \gamma$ and $\lim_{|\mathbf{k}|\to\infty} P(f|\mathbf{v_k}) = (1-p) - \gamma$.

Extra signals do nothing but introduce noise. They reduce the chance of a correct belief, but they also reduce the chance of a mistaken belief. In the extreme case, they wash out and provide negligible information.

Corollary 1. For small $\varepsilon > 0$, let $p = .5 + \varepsilon$ and $\gamma = 1 - p$. Then $\lim_{|\mathbf{k}|\to\infty} P(e|\mathbf{v_k}) = 1 - 2\varepsilon$.

Even though each signal is different and informative, their dependence creates a sea of noise. In contrast to the CJT, an increasingly large group is no closer to learning the truth.

Now suppose the group ignores any dependence, as most of the literature assumes. Instead, members treat signals as if they were independent by taking a majority vote. As it turns out, a group vote is then no different than a single vote. That is,

Proposition 3. $\forall \mathbf{k} \in \mathbb{K}$ s.t. $|\mathbf{k}| \in \{2x-1|x \in \mathbb{N}^+\}$, (S2) and (A1)-(A3) $\Rightarrow$
(1) $P(c|\mathbf{v_k}) = p$ and $P(f|\mathbf{v_k}) = (1-p)$, and
(2) $\lim_{|\mathbf{k}|\to\infty} P(c|\mathbf{v_k}) = p$, and $\lim_{|\mathbf{k}|\to\infty} P(f|\mathbf{v_k}) = (1-p)$.

The marginal group member provides zero marginal benefit. For any size group, the probabilities of a false and correct decision are simply those of a single member.[5]

For the class of dependent signals given by (A3), expanding membership adds noise at worst and nothing at best. Interestingly, the result holds whether or not the group accounts for statistical dependence.

IV.2. **Experts.** If experts have more information, they should make better decisions on average. In the model, a better average decision comes through a higher chance of a correct signal, p_i. Although p_i is greater than one half for every individual by assumption, it can be helpful to refer to those with a relatively high p_i as "experts" and those with a relatively low p_i as "inexperts." Intuitively, a group of experts should make better decisions on average than a group of inexperts. Showing the result in terms of *Proposition 1* is an easy exercise. For any group size, both the probability of a correct decision and the rate of its convergence increase in p.

Table 1

$\mathbf{s}$	$P(\mathbf{s}\|\mathcal{A})$	$P(\mathbf{s}\|\mathcal{B})$
$a_1 \cap a_2$	x	$1-p_1-y$
$a_1 \cap b_2$	p_1-x	y
$b_1 \cap a_2$	y	p_1-x
$b_1 \cap b_2$	$1-p_1-y$	x

Less obvious is what benefit a single expert can bring to a group of inexperts. Does the expert raise up the group? Does the group bring down the expert? Or is it some combination of the two? *Proposition 4* gives an optimistic answer. Suppose the group accounts for statistical dependence in its beliefs. In terms of believing in the wrong option, the group can never do worse than its best member.[6]

Proposition 4. (S1), (A1), and (A2) $\Rightarrow P(f|\mathbf{v_n}) \leq 1 - \max\{p_1, p_2, ..., p_n\}\ \forall n \in \mathbb{N}^+$, $g \in \mathcal{G}_n$.

[5] Note that *Proposition 3* does require n to be odd. As discussed in Section II, allowing an even n with a coin flip tie-breaking rule only introduces extra cases without altering the main conclusion. Asymptotically, the parity of n is unimportant because the chance of a tie vote converges to zero.

[6] Henry David Thoreau once quipped, "The mass never comes up to the standard of its best member, but on the contrary degrades itself to a level with the lowest" (Thoreau and Shepard 1961, at 4). *Proposition 4* shows just the opposite—the best member elevates the lowest.

The power in the result is its generality. *Proposition 4* holds under arbitrary dependence. A similar result applies to the probability of an uninformative outcome. The factor of two here appears because for symmetric signals, any uninformative event must come with a pair.

Proposition 5. (S1), (A1), and (A2) $\Rightarrow P(e|\mathbf{v_n}) \leq 2(1-\max\{p_1, p_2, ..., p_n\})\ \forall n \in \mathbb{N}^+,\ g \in \mathcal{G}_n$.

An expert is therefore an insurance policy against bad or uninformed choices. For example, if an expert has $p_i = 9/10$, the chance of the group choosing the wrong outcome is no greater than 1/10, and the chance of deriving no information from communication is no greater than 2/10. Less fortunately, the bounds on bad outcomes are tight. The case of *Proposition 4* is trivial for $n = 1$. For the bound on *Proposition 5*, consider the general case for $n = 2$ presented by Table 1. Let $x + y = p_1$ and $x = 2p_1 - 1$. Then $P(e|\mathbf{v_2}) = p_1 - x + y = 2(1 - p_1) = 2(1 - \max\{p_1, p_2\})$.

IV.3. **When More is Worse.** For an expert to bound mistaken beliefs, the group must update rationally. Unfortunately, an expert may be powerless if the group ignores dependence. Take the intuitive case of a single expert amid a less informed group. The expert has accuracy $\bar{p}$, and the other n members have a lower $\underline{p}$. Denoting the expert's unrealized signal by v_0,

Claim 1. $\forall\, 1 > \bar{p} > \underline{p} > 1/2,\ \exists n \in \mathbb{N}^+, g \in \mathcal{G}_n$, s.t. $P(c|v_0, \mathbf{v_n}) < \bar{p}$, and $P(f|v_0, \mathbf{v_n}) > (1-\bar{p})$.

Since positively correlated signals contain less information than independent ones, they are overweighted when votes are counted equally. Thus, if the lesser informed contingent has enough positively correlated votes, their voice will drown out the expert, who would otherwise make better decisions than the group.[7]

In summary, expertise is everywhere beneficial, and group size cannot always compensate for its absence. Whether or not a group accounts for statistical dependence, the signals may combine to create a sea of noise, providing no information. If the group does account for dependence, an expert always raises up the other members, but if not, the other members can bring an expert down.

V. Social Welfare

Since the CJT does not hold under arbitrary statistical dependence, increasing participation in a group vote is no guarantee of a sound decision. On the other hand, recruiting an expert does provide some guarantee (see *Propositions 4* and *5*). The tradeoff between expertise and participation therefore raises the question of social welfare. Returning to the classical independence setting, consider the optimal group size when participation is costly. Simple intuition is unclear as to whether participation should rise with expertise (as measured by $p = p_i\, \forall i \in \mathbf{n}$). Experts add more to the group than inexperts, but a group might try to compensate for a lack of expertise by increasing membership.

[7]This result complements Glaeser and Sunstein (2009), who present a similar conclusion with learning over a continuous, normal distribution.

Suppose a population of size N is choosing whether to adopt a new policy. Let $d = 1$ denote a correct decision, i.e., instituting the policy when and only when it improves welfare, and $d = 0$ an incorrect decision. The vote yields a collective benefit $u(d, N) = duN$. If the policy is a public good, then u is the benefit per individual of the better outcome, with the utility of the worse outcome normalized to zero. Expertise is homogenous across the population, so the opportunity cost of participation, w, is constant. If the group recruits n members of expertise p, the net expected utility of a vote is

$$EU(p, n) = F(p, n)uN - nw, \tag{1}$$

where $F(p, n)$ is the probability of a correct decision given p and n, as defined in Section II.

V.1. **Optimal Participation.** For a given level of expertise, a group optimizes its membership by choosing n to maximize equation (1).[8] Let $n^*(p) \in \{1, 2, ..., N\}$ denote the solution. Since the choice set is finite and expected utility is bounded, $n^*(p)$ is well defined. Assume $F(p, 1)uN > \gamma_0 + \gamma_1(p)$ to avoid the trivial solution of zero membership. Then optimal participation rises and falls in p. (But since n is discrete, $n^*(p)$ might be everywhere decreasing for certain utility parameters.) Expertise and participation are complements for low p and substitutes for high p.

Proposition 6. The function $n^*(p)$ has a single peak.

Intuitively, a good decision requires a modicum of expertise. If a group is uninformed, participation is simply not worth the cost. On the other hand, if a group is sufficiently informed, then participation and expertise become substitutes. The marginal benefit of an expert to the decision will eventually exceed her marginal cost of participation. Table 2 provides a demonstration for a policy of relatively small importance, with $u - w = 1$.

Table 2. Optimal Participation

p	.51	.53	.55	.6	.7	.8	.9
$n^*(p)$	27	149	153	85	33	17	9
$F(p, n^*(p))$	0.541	0.769	0.893	0.969	0.992	0.997	0.999

Note: $N = 999$, $u = 4$, $\gamma_0 = 3$, $\gamma_1(p) = 0$.

According to the CJT, group size compensates for low expertise, but opportunity costs present a normative obstacle. Both very uninformed and very informed groups do best by assigning a choice to a small committee. Of course, for a policy of great importance, even an uninformed population should devote its every resource to a resolution.

[8]This formulation of the problem abstracts from the incentive to free ride on the participation of others. Some real world settings eliminate the incentive, as in jury duty and other organizations with committee service requirements.

V.2. **Optimal Investment.** The next logical step is to ask how much a group should invest in creating expertise. Assume expertise $p \in [1/2, 1]$ can be created at expense $\phi(p)$ per member, with $\phi'(p) > 0$, $\phi''(p) >$, $\phi(1/2) = 0$, and $\lim_{p \to 1} \phi(p) = \infty$. For simplicity, suppose the group invests equally in each member. For a given group size, the optimal investment maximizes

$$EU(p, n) = F(p, n)uN - n(w + \phi(p)). \tag{2}$$

As before, w is the opportunity cost of participation. Since the cost of investment is convex, the following lemma implies the optimal investment is well defined.

Lemma 4. $F(p, n)$ is strictly concave in $p \in [1/2, 1]$.

The optimal group size then maximizes (2) given the investment rule. Again, with a finite objective function over a finite choice set, the problem is well defined. Writing the optimal investment as a function of group size,

Lemma 5. $\dfrac{\partial p^*(n)}{\Delta n} < 0$.

Society reduces the investment per individual when it invests in a larger proportion of the population. Thus, *Lemma 5* implies an ambiguous relationship between welfare and participation. As participation rises, social cost rises directly through the opportunity cost but falls indirectly through the lower investment per individual. The probability of a correct decision also rises directly through participation but falls indirectly through the lower investment per individual. The net impact on expected utility is therefore indeterminate.

TABLE 3. Optimal Participation with Investment

$c(p)$	$\frac{1}{(1-p)} - 2$		$\frac{p^4}{(1-p)} - (1/2)^3$		$tan((\pi - 1)(p/2) + 1/2) - tan((\pi - 1)(1/4) + 1/2)$	
w	3	1	3	1	3	1
u	4	10,000	4	10,000	4	10,000
n^*	5	29	7	37	5	27
$p^*(n^*)$	0.916	0.861	.889	.834	.918	.869

Note: N=999 for all cases.

Depending on the parameters, an inclusive education policy may or may not be desirable, but consider the examples in Table 3. The optimality of universal education implies a grossly implausible value function. Results are quite robust to various specifications. For example, if $N = 999$ and $\phi = p^4/(1-p) - (1/2)^3$, a u/w ratio of one billion to one implies an optimal participation rate of only 8%. If the investment can apply broadly to numerous decisions, participation rates climb slightly higher, but elitism remains the better policy. Thus, even in the "best-case" scenario of an independent, homogenous population, society should invest in a group of elite members and rely on their expertise. The reality of variable training costs among individuals only amplifies the result, as does the possibility of harmful signal correlations.

Although the functions in Table 3 share an asymptote at $p = 1$, many applications might well have an earlier one. Van Such et al. (2017) compile a sample of patients seeking visiting doctors for second opinions, finding only 12% received a complete and correct diagnosis the first time, while 21% received a completely wrong diagnosis. Training standards for doctors are among the highest of any profession, so achieving diagnostic reliability of p near 1 seems unattainable with current technology and patient loads.[9] However, imposing a smaller asymptote generally fails to reverse the superiority of elitism. For example, if $N = 999$, $\phi = 1/(4/5 - p) - (10/3)^3$, and $u/w = 10,000$, the asymptote at $p = .8$ raises optimal participation to $n^* = 127$.

VI. An Optimistic Finite Result

To this point, every conclusion favors elitism over democracy. Since relative inexperts can also negate the value of an expert if they ignore statistical dependence, delegation of a decision to an expert seems all the more attractive (see *Claim 1*). This section provides a simple yet powerful counterpoint. If properly acknowledged, statistical dependence can be an asset.

Table 4. Uncovering the True State with Certainty

	$\mathcal{A}$	$\mathcal{B}$
$a_1 \cap a_2 \cap a_3$	$p - x$	$1-p-2x$
$a_1 \cap a_2 \cap b_3$	0	x
$a_1 \cap b_2 \cap a_3$	0	x
$a_1 \cap b_2 \cap b_3$	x	0
$b_1 \cap b_2 \cap b_3$	$1-p-2x$	$p - x$
$b_1 \cap b_2 \cap a_3$	x	0
$b_1 \cap a_2 \cap b_3$	x	0
$b_1 \cap a_2 \cap a_3$	0	x

Note: If $x = (1/2)(1-p)$, the agent learns the true state for $p \geq 1/3$.

Consider the signals in Table 4 for a group of size three. The group uncovers the state of the world with certainty if $x = (1/2)(1 - p)$. The case of $p = 1/2$ is illustrative. Taken on their own, the signals amount to uninformative coin flips, but taken altogether, they perfectly reveal the true state of the world. For $1/3 \leq p < 1/2$, even a group of dunces can learn the true state if they are clever enough to exploit the dependence in their signals.[10]

[9] A diagnosis is not a binary decision, but for the sake of argument, consider the question of whether or not a patient has a particular ailment. Some diseases are notoriously hard to diagnose.

[10] Three is the minimal number of signals required to uncover the true state. To see why, consider the general case for $n = 2$ in Table 1. Assume learning is complete. If $x = y = 0$, then $P(a_2|\mathcal{A}) = 0$, a contradiction. If $y > 0$, then $x = 0$, which implies $y = 1-p$, but then $P(a_2|\mathcal{A}) = P(a_2 \cap a_1|\mathcal{A}) + P(a_2 \cap b_1|\mathcal{A}) = 0 + 1 - p < 1/2$, again a contradiction. Lastly, if $x > 0$, then $y = 0$, which implies $x = p$, but then both signals are identical.

Claim 2. For the signals in Table 1, learning is complete if $x = (1/2)(1 - p)$ and $p \geq 1/3$.

If the group wrongly believes in the independence assumption of the Condorcet world, they will conclude their information has zero value, when in fact their information is perfect. While independence is often treated as the ideal scenario, *Claim 2* upends the intuition. The CJT requires an asymptotically large group of informed members to achieve certainty. The right form of dependence requires only three members, and their signals need not be informative.

VII. Conclusion

A population can obviate expertise if it can only identify or create the right statistical dependencies among its members. With no obvious mechanism available for meeting the goal, however, countries and organizations face difficult choices. Should they foster democratic participation? Should they invest in universal education? In theory, such egalitarian policies are inferior to elitism. Unacknowledged statistical dependencies can yield poor democratic decisions, and less informed members of the population can drown out experts. Furthermore, even in a homogenous population of statistically independent members, elitism is the better investment strategy for all but the most extreme utility parameters.

The results above assume symmetry in the signals, but perhaps asymmetry could offer another point of optimism. That point is left for future research, as the topic remains largely unexplored in the literature. In a recent exception, Stone (2015) presents a model in which some members have a more accurate signal in one state of the world, while the remaining members have a more accurate signal in the opposite state. Depending on the parameters, the asymmetry can be good or bad for the committee.

The results may also have use beyond the question of democracy. Economists frequently model Bayesian learning through binary signals. For example, Bikhchandani et al. (1992) examine how individuals react to learning the votes of others before them in sequence, which partly inspired a literature on herding and information cascades (cf. Banerjee 1992). Rabin and Schrag (1999) model a process of confirmation bias through a repeated binary signal. Introducing differential expertise into similar applications might yield further insight on its importance to good decisions.

References

Ahn, David S., and Santiago Oliveros. 2014. The Condorcet Jur(ies) Theorem. *Journal of Economic Theory* 150: 841–851.

Austen-Smith, David, and Jeffrey S. Banks. 1996. Information aggregation, rationality, and the Condorcet jury theorem. *The American Political Science Review* 90(1): 34–45.

Banerjee, Abhijit V. 1992. A simple model of herd behavior. *The Quarterly Journal of Economics* 107(3): 797–817.

Bénabou, Roland. 2013. Groupthink: Collective delusions in organizations and markets. *Review of Economic Studies* 80(2): 429–462.

Ben-Yashar, Ruth. 2006. Information is important to Condorcet jurors. *Public Choice* 127(3/4): 313–327.

Ben-Yashar, Ruth. 2014. The generalized homogeneity assumption and the Condorcet jury theorem. *Theory and Decision* 77(2): 237–241.

Ben-Yashar, Ruth, and Jacob Paroush. 2000. A nonasymptotic Condorcet Jury Theorem. *Social Choice and Welfare* 17(2): 189–199.

Ben-Yashar, Ruth, and Mor Zahavi. 2011. The Condorcet jury theorem and extension of the franchise with rationally ignorant voters. *Public Choice* 148(3/4): 435–443.

Berend, Daniel, and Jacob Paroush. 1998. A nonasymptotic Condorcet Jury Theorem. *Social Choice and Welfare* 15(4): 481–488.

Berend, Daniel, and Luba Sapir. 2007. Monotonicity in Condorcet's Jury Theorem with dependent voters. *Social Choice and Welfare* 28(3): 507–528.

Berg, Sven. 1993. Condorcet's jury theorem, dependency among jurors. *Social Choice and Welfare* 10(1): 71–83.

Berg, Sven. 1994. Evaluation of some weighted majority decision rules under dependent voting. *Mathematical Social Sciences* 28(2): 71–83.

Bikhchandani, Sushil, Hirshleifer, David, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *The Journal of Political Economy* 100(5): 992–1026.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, Dvid McKenzie, and John Roberts. 2013. Does management matter? Evidence from India. *The Quarterly Journal of Economics* 128(1): 1–51.

Boland, Philip J. 1989. Majority systems and the Condorcet Jury Theorem. *Journal of the Royal Statistical Society* 38(3): 181–189.

Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro. 2015. Do pharmacists buy Bayer? Informed shoppers and the brand premium. *The Quarterly Journal of Economics* 130(4): 1669–1726.

Charness, Gary, and Matthias Sutter. 2012. Groups make better self-interested decisions. *Journal of Economic Perspectives* 26(3): 157–176.

Condorcet, M. (1785). *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris. See I. McLean and F. Hewitt, trans., 1994.

Congleton, Roger D. 2007. Informational limits to democratic public policy: The jury theorem, yardstick competition, and ignorance. *Public Choice* 132(3/4): 333–352.

Coughlan, Peter J. 2000. In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *The American Political Science Review* 94(2): 375–393.

Crawford, Vincent P., and Joel Sobel. 1982. Strategic information transmission. *Econometrica* 50(6): 1431–1451.

DeGroot, Morris H. 1974. Reaching a consensus. *Journal of the American Statistical Association* 69 (345): 118–121.

Dharmapala, Dhammika, and Richard H. McAdams. 2003. The Condorcet Jury Theorem and the expressive function of the law: A theory of informative law. *The American Law and Economics Review* 5(1): 1–31.

Feddersen, Timothy, and Wolfgang Pesendorfer. 1998. The inferiority of unanimous jury verdicts under strategic voting. *The American Political Science Review* 92(1): 23–35.

Gerardi, Dino, and Leeat Yariv. 2008. Costly expertise. *American Economic Review: Papers & Proceedings* 98(2): 187–193.

Glaeser, Edward L., and Cass R. Sunstein. 2009. Extremism and social learning. *Journal of Legal Analysis* 1(1): 263–324.

Hilger, Nathaniel G. 2016. Why don't people trust experts. *Journal of Law & Economics* 59(2): 293–311.

Jackson, Matthew O., Brian W. Rogers, and Yves Zenou. 2017. The economic consequences of social-network structure. *Journal of Economic Literature* 55(1): 49–95.

Kaniovski, Serguei. 2010. Aggregation of correlated votes and Condorcet's Jury Theorem. *Theory and Decision* 69(3): 453–468.

Katzner, Donald W. 1995. Participatory decision-making in the firm. *Journal of Economic Behavior and Organization* 26(2): 221–236.

Koriyama, Yukio, and Balázs Szentes. 2009. A resurrection of the Condorcet Jury Theorem. *Theoretical Economics* 4: 227–252.

Krishna, Vijay, and John Morgan. 2001. A model of expertise. *The Quarterly Journal of Economics* 116(2): 747–775.

Ladha, Krishna K. 1992. The Condorcet Jury Theorem, free speech, and correlated votes. *American Journal of Political Science* 36(3): 617–634.

Ladha, Krishna K. 1993. Condorcet's jury theorem in light of de Finett's theorem: Majority-rule voting withcorrelated votes. *Social Choice and Welfare* 10(1): 69–85.

Ladha, Krishna K. 1995. Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. *Journal of Economic Behavior and Organization* 26(3): 353–372.

Laslier, Jean-François, and Jörgen W. Weibull. 2011. An incentive-compatible Condorcet Jury Theorem. *Scandinavian Journal of Economics* 115(1): 84–108.

Levy, Gilat, and Ronny Razin. 2015. Correlation neglect, voting behavior, and information aggregation. *American Economic Review* 105(4): 1634–1645.

Li, Hao, and Wing Suen. 2004. Delegating decisions to experts. *Journal of Political Economic* 112(1): 311–335.

Lindner, Ines. 2008. A generalization of Condorcet's Jury Theorem to weighted voting games with many small voters. *Economic Theory* 35(3): 607–611.

McCannon, Bryan C., and Paul Walker. 2016. Endogenous competence and a limit to the Condorcet Jury Theorem. *Public Choice* 169(1): 1–18.

McLennan, Andrew. 1998. Consequences of the Condorcet Jury Theorem for beneficial information aggregation by rational agents. *The American Political Science Review* 92(2): 413–418.

McMurray, Joseph C. 2013. Aggregating information by voting: The wisdom of the experts versus the wisdom of the masses. *Review of Economic Studies* 80(1): 277–312.

Meirowitz, Adam. 2002. Informative voting and Condorcet Jury Theorems with a continuum of types. *Social Choice and Welfare* 19(1): 219–236.

Milgrom, Paul, and Chris Shannon. 1994. Monotone comparative statics. *Econometrica* 62(1): 157–180.

Miller, N. R. 1986. Information, electorates, and democracy: Some extensions and interpretations of the Condorcet Jury Theorem. In: Grofman B, Owen G (eds.) Information pooling and group decision making. Jai, Greenwich, CT.

Mukhopadhaya, Kaushik. 2003. Jury size and the free rider problem. *Journal of Law, Economics, & Organization* 19(1): 24–44.

Myerson, Roger B. 1998. Extended Poisson games and the Condorcet Jury Theorem. *Games and Economic Behavior* 25(1): 111–131.

Nitzan, Shmuel, and Jacob Paroush. 1984. The significance of independent decisions in uncertain dichotomous choice situations. *Theory and Decision* 17(1): 47–60.

Paroush, Jacob. 1998. Stay away from fair coins: A Condorcet Jury Theorem. *Social Choice and Welfare* 15(1): 15–20.

Pelig, Bezalel, and Shmuel Zamir. 2012. Extending the Condorcet Jury Theorem to a general dependent jury. *Social Choice and Welfare* 39(1): 91–125.

Persico, Nicola. 2004. Committee design with endogenous information. *Review of Economic Studies* 71(1): 165–191.

Rabin, Matthew, and Joel L. Schrag. 1999. First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics* 114(1): 37–82.

Sah, Raaj Kumar, and Joseph E. Stiglitz. 1988. Committees, hierarchies, and polyarchies. *The Economic Journal* 98(391): 451–470.

Stone, Peter. 2015. Introducing difference into the Condorcet Jury Theorem. *Theory and Decision* 78(3): 399–409.

Sunstein, Cass R. 2005. Group judgments: Statistical means, deliberation, and information markets. *New York University Law Review* 80(3): 962–1049.

Surowiecki, James. 2005. *The wisdom of crowds*. Anchor Books.

Thoreau, Henry David, and Odell Shepard. 1961. *The Heart of Thoreau's Journals*. Dover Publications (1st Edition).

Van Such, Monica, Robert Lohr, Thomas Beckman, and James M. Naessens. 2017. Extent of diagnostic agreement among medical referrals. *Journal of Evaluation in Clinical Practice* 23(4): 870–874.

Wit, Jörgen. 1998. Rational choice and the Condorcet Jury Theorem. *Games and Economic Behavior* 22: 364–376.

Appendix

Proof of Proposition 1. See Boland (1989).

Proof of Lemma 1. Assume (A1)-(A3). Take any $\mathbf{k} \in \mathbb{K}$ with $|\mathbf{k}| = \infty$. For $i \in \mathbf{k}$, define $\sigma_i = s_{k_i}$, $\alpha_i = a_{k_i}$, and $\beta_i = b_{k_i}$. Let $\delta_1 = \sum_{j=1}^{k_1-1} d_j$ and $\delta_{i+1} = \sum_{j=1}^{k_{i+1}-1} d_j - \delta_i$. By construction,

$$\sum_{\substack{j=1\\(j\in\mathbf{k})}}^{m-1} \delta_j = \sum_{j=1}^{m-1} d_j \Rightarrow \sum_{\substack{j=l\\(j\in\mathbf{k})}}^{m-1} \delta_j = \sum_{j=1}^{m-1} d_j \;\; \forall l, m \in \mathbb{N}^+,\; l < m,\; \text{with } (m-1) \in \mathbf{k}.$$

Define $f(\cdot)$ as the ordered bijection between $\mathbb{N}^+$ and $\mathbf{k}$. That is, $f(1) = k_1, f(2) = k_2$, and so on. (Since both $\mathbb{N}^+$ and $\mathbf{k}$ are countably infinite, such a bijection exists.) Let $\mathbf{k}' \subset \mathbf{k}$ with $|\mathbf{k}'| \notin \{\emptyset, \infty\}$. For $m \in \mathbb{N}^+$, define $m' = f(m)$. Then

$$P(\alpha_{k'_{|\mathbf{k}'|}+m'} \cap (\bigcup_{i\in\mathbf{k}'} \alpha_i)|\mathcal{A}) = P(a_{k'_{|\mathbf{k}'|}+m'} \cap (\bigcup_{i\in\mathbf{k}'} a_i)|\mathcal{A}) = p - \sum_{j=k'_{|\mathbf{k}'|}}^{k'_{|\mathbf{k}'|}+m'-1} d_j = p - \sum_{\substack{j=k'_{|\mathbf{k}'|}\\ j\in\mathbf{k}}}^{k'_{|\mathbf{k}'|}+m'-1} \delta_j.$$

Similarly,

$$P(\beta_{k'_{|\mathbf{k}'|}+m'} \cap (\bigcap_{i\in\mathbf{k}'} \alpha_i)|\mathcal{A}) = P(b_{k'_{|\mathbf{k}'|}+m'} \cap (\bigcap_{i\in\mathbf{k}'} a_i)|\mathcal{A}) = \sum_{j=k'_{|\mathbf{k}'|}}^{k'_{|\mathbf{k}'|}+m'-1} d_j = \sum_{\substack{j=k'_{|\mathbf{k}'|}\\ j\in\mathbf{k}}}^{k'_{|\mathbf{k}'|}+m'-1} \delta_j.$$

Lastly,

$$\sum_{\substack{j=1\\(j\in\mathbf{k})}}^{\infty} \delta_j = \sum_{j=1}^{\infty} d_j \leq 1 - p$$

by construction. Thus, (A2) is recreated for the analogous sequence $\delta_i, \sigma_i \in \{\alpha_i, \beta_i\}$ under the bijection $f(\cdot)$. □

Proof of Lemma 2. Assume (A1)-(A3). We seek to show

$$P(\mathbf{s_k}|\mathcal{A}) = P(\bar{\mathbf{s}}_\mathbf{k}|\mathcal{A}) \,\forall\, \mathbf{s_k} \in S_\mathbf{k} \text{ s.t. } \exists\, i, j \in \mathbf{k} \text{ for which } s_i \neq s_j. \tag{L2$'$}$$

By the symmetry assumption, (L2$'$) implies $P(\mathbf{s_k}|\mathcal{A}) = P(\mathbf{s_k}|\mathcal{B})$.

The proof uses the following two results. First, any sequence of events containing a "b" signal between two "a" signals or vice versa must have probability zero. Suppressing the conditional probability notation,

$$P(\mathbf{s_k}) = 0 \;\text{ if }\; \exists\, i < j < k \text{ for which } s_i \neq s_j \neq s_k. \tag{L2.A}$$

Second, defining $\mathbf{k}_j^+ = \{i \in \mathbf{k} | i \geq j\}$ and $\mathbf{k}_j^- = \{i \in \mathbf{k} | i < j\}$ for $j \in \mathbf{k}/\{k_1\}$,

$$\begin{aligned} P(\mathbf{s_k}) = \delta_{j-1} \;\text{ if }\; \exists j \in \mathbf{k} \;\text{ s.t. } & (s_i = a_i \,\forall i \in \mathbf{k}_j^-) \cap (s_i = b_i \,\forall i \in \mathbf{k}_j^+) \\ \text{or } & (s_i = b_i \,\forall i \in \mathbf{k}_j^-) \cap (s_i = a_i \,\forall i \in \mathbf{k}_j^+), \end{aligned} \tag{L2.B}$$

where δ_{j-1} is constructed per Lemma 1. Since (L2′) implies at least two of the signals are different, (L2.A) and (L2.B) cover all the relevant cases. Suppose $\exists\, i < j < k$ for which $s_i \neq s_j \neq s_k$. Then $P(\mathbf{s_k}) = 0$ by (L2.A). But $s_i \neq s_j \neq s_k \Rightarrow \bar{s}_i \neq \bar{s}_j \neq \bar{s}_k$. Then $P(\bar{\mathbf{s}}_\mathbf{k}) = 0$ by (L2.A) again, so $P(\mathbf{s_k}) = P(\bar{\mathbf{s}}_\mathbf{k})$. For the remaining sequences in the class of (L2.B), the result is immediate. □

Proof of L2.A. Suppose $s_i = a_i$, $s_j = b_j$, and $s_k = a_k$ (the proof for the opposite case is similar). Per Lemma 1, define $\alpha_1 = a_i$, $\alpha_2 = a_j$, $\alpha_3 = a_k$, and $\beta_1 = b_i$, $\beta_2 = b_j$, $\beta_3 = b_k$. Also, define $\delta_1 = \sum_{m=1}^{i-1} d_m$, $\delta_2 = \sum_{m=1}^{j-1} d_m - \delta_1$, and $\delta_3 = \sum_{m=1}^{k-1} d_m - \delta_2$. By the application of (A2) to the redefined system,

$$\begin{aligned}
&P(\alpha_1 \cap \alpha_2) = p([(\alpha_1 \cap \alpha_2) \cap \alpha_3] \cup [(\alpha_1 \cap \alpha_2) \cap \beta_3]) = P(\alpha_1 \cap \alpha_2 \cap \alpha_3) + P(\alpha_1 \cap \alpha_2 \cap \beta_3)\\
&\Rightarrow p - \delta_1 = P(\alpha_1 \cap \alpha_2 \cap \alpha_3) + \delta_2\\
&\Rightarrow P(\alpha_1 \cap \alpha_2 \cap \alpha_3) = p - \delta_1 - \delta_2,
\end{aligned}$$

where the first step makes use of the complementarity of a_3 and b_3. Then

$$\begin{aligned}
&P(\alpha_1 \cap \alpha_3) = P(\alpha_1 \cap \alpha_2 \cap \alpha_3) + P(\alpha_1 \cap \beta_2 \cap \alpha_3)\\
&\Rightarrow p - \delta_1 - \delta_2 = p - \delta_1 - \delta_2 + P(\alpha_1 \cap \beta_2 \cap \alpha_3)\\
&\Rightarrow P(\alpha_1 \cap \beta_2 \cap \alpha_3) = 0. \ \square
\end{aligned}$$

Proof of L2.B. The proof is by induction. Apply Lemma 1 to the index $\mathbf{k+1}$ nesting $\mathbf{k}$ (with $|\mathbf{k+1}| = |\mathbf{k}| + 1$). That is, map $\{k_1, k_2, ..., k_{|\mathbf{k}|}, k_{|\mathbf{k+1}|}\}$ to $\{1, 2, ..., |\mathbf{k}|, |\mathbf{k+1}|\}$. First, by the application of (A2) to the redefined system, $P(\alpha_1 \cap \beta_2) = \delta_1$. Then

$$\begin{aligned}
P(\beta_1 \cap \alpha_2) = 1 - P(\overline{\beta_1 \cap \alpha_2}) = 1 - P(\alpha_1 \cup \beta_2) &= 1 - [P(\alpha_1) + P(\beta_2) - P(\alpha_1 \cap \beta_2)]\\
&= 1 - [p + 1 - p - \delta_1]\\
&= \delta_1,
\end{aligned}$$

where the second equality follows from De Morgan's Law, so the induction hypothesis holds for $|\mathbf{k}| = 2$. For $|\mathbf{k}|+1$, define $\mathbf{k}_j^+ = \{i | j \leq i \leq |\mathbf{k}|\}$ and $\mathbf{k}_j^- = \{i | 1 \leq i < j\}$ for $j \in \mathbf{k}$. Assume $\sigma_i = \alpha_i \,\forall i \in \mathbf{k}_j^-$ and $\sigma_i = \beta_i \,\forall i \in \mathbf{k}_j^+$ (the proof for the opposite case is similar). That is, consider a sequence of the form $(\alpha_1, \alpha_2, ..., \alpha_{j-1}, \beta_j, \beta_{j+1}, ..., \beta_{|\mathbf{k}|})$ for some $j \in \{2, 3, ..., |\mathbf{k}|\}$. Then

$$\begin{aligned}
P(\alpha_1, \alpha_2, ..., \alpha_{j-1}, \beta_j, \beta_{j+1}, ..., \beta_{|\mathbf{k}|}) &= P(\alpha_1, \alpha_2, ..., \alpha_{j-1}, \beta_j, \beta_{j+1}, ..., \beta_{|\mathbf{k}|}, \beta_{|\mathbf{k}|+1})\\
&\quad + P(\alpha_1, \alpha_2, ..., \alpha_{j-1}, \beta_j, \beta_{j+1}, ..., \beta_{|\mathbf{k}|}, \alpha_{|\mathbf{k}|+1})\\
&= P(\alpha_1, \alpha_2, ..., \alpha_{j-1}, \beta_j, \beta_{j+1}, ..., \beta_{|\mathbf{k}|}, \beta_{|\mathbf{k}|+1})\\
&\quad + 0
\end{aligned}$$

by (L2.A), so

$$P(\alpha_1, \alpha_2, ..., \alpha_{j-1}, \beta_j, \beta_{j+1}, ..., \beta_{|\mathbf{k}|}, \beta_{|\mathbf{k}|+1}) = P(\alpha_1, \alpha_2, ..., \alpha_{j-1}, \beta_j, \beta_{j+1}, ..., \beta_{|\mathbf{k}|}) = \delta_{j-1}. \ \square$$

Proof of Proposition 2. Let $\mathbf{y} \subseteq \mathbf{k} \in \mathbb{K}$. Assume (S1), (A1)-(A3), and $\mathcal{S} = \mathcal{A}$ (the proof for $\mathcal{S} = \mathcal{B}$ is similar). Suppressing the conditional notation for the state of the world,

$$P(c|\mathbf{v_k}) = P(\underset{i\in\mathbf{k}}{\cap} v_i = a_i) \leq P(\underset{i\in\mathbf{k}\backslash\mathbf{y}}{\cap} v_i = a_i) = P(c|v_{\mathbf{k}\backslash\mathbf{y}}),$$

since $\mathbf{k}\backslash\mathbf{y} \subseteq \mathbf{k}$. The equalities follow from Lemma 2. Similarly,

$$P(f|\mathbf{v_k}) = P(\underset{i\in\mathbf{k}}{\cap} v_i = b_i) \leq P(\underset{i\in\mathbf{k}\backslash\mathbf{y}}{\cap} v_i = b_i) = P(c|v_{\mathbf{k}\backslash\mathbf{y}}),$$

which establishes (1). For (2), note that (A3) implies $P(c|\mathbf{v_k}) = p - \sum_{j=1}^{|\mathbf{k}|-1} \delta_j$ by way of Lemma 1. Likewise, $P(f|\mathbf{v_k}) = (1-p) - \sum_{j=1}^{|\mathbf{k}|-1} \delta_j$. Then

$$\lim_{|\mathbf{k}|\to\infty} P(c|\mathbf{v_k}) = \lim_{n\to\infty} p - \sum_{j=1}^{n} \delta_j = p - \gamma, \text{ and}$$

$$\lim_{|\mathbf{k}|\to\infty} P(f|\mathbf{v_k}) = \lim_{n\to\infty} (1-p) - \sum_{j=1}^{n} \delta_j = (1-p) - \gamma. \ \square$$

Proof of Proposition 3. Take $\mathbf{k} \in \mathbb{K}$ s.t. $|\mathbf{k}| \in \{2x-1|x \in \mathbb{N}^+\}$. Assume (S2), (A1)-(A3), and $\mathcal{S} = A$ (the proof for $\mathcal{S} = \mathcal{B}$ is similar). The agent believes in the correct state of the world if and only if she receives more "a" than "b" signals. Suppressing the conditional notation for the state of the world, (A3) implies $P(\cap_{i\in\mathbf{k}} v_i = a_i) = p - \sum_{j=1}^{k_{|\mathbf{k}|}-1} d_j$ and $P(\cap_{i\in\mathbf{k}} v_i = b_i) = (1-p) - \sum_{j=1}^{k_{|\mathbf{k}|}-1} d_j$. Define $\mathbf{k}_j^+ = \{i \in \mathbf{k}|i \geq j\}$ and $\mathbf{k}_j^- = \{i \in \mathbf{k}|i < j\}$ for $j \in \mathbf{k}/\{k_1\}$. By Lemma 2, $P(\mathbf{s_k}) = 0$ if $\exists i < j < k$ for which $s_i \neq s_j \neq s_k$. Also,

$$\begin{aligned} P(\mathbf{s_k}) = \delta_{j-1} \text{ if } \exists j \in \mathbf{k} \text{ s.t. } & (s_i = a_i \, \forall i \in \mathbf{k}_j^-) \cap (s_i = b_i \, \forall i \in \mathbf{k}_j^+) \\ \text{or } & (s_i = b_i \, \forall i \in \mathbf{k}_j^-) \cap (s_i = a_i \, \forall i \in \mathbf{k}_j^+). \end{aligned}$$

Therefore, excluding the cases where every signal is identical, the agent is equally likely to receive more "a" or "b" signals. Furthermore, since $|\mathbf{k}|$ is odd, the probability of getting an equal number of each type of signal is zero, implying $P(c|\mathbf{v_k}) = p$ and $P(f|\mathbf{v_k}) = 1-p$. Since the equalities hold for all finite $|\mathbf{k}|$, they also hold in the limit. $\square$

Proof of Claim 1. Assume the lesser informed members share a perfectly correlated signal. Then with m lesser informed members, $a_1 = a_1 \cap a_2 \cap ... \cap a_m$. Although perfect correlation violates (A1), introducing miniscule differences in the signals yields the same conclusion, only messier. Let $\mathcal{S} = \mathcal{A}$ (the proof for $\mathcal{S} = \mathcal{B}$ is similar). Under (S2), Bayes' rule gives

$$P(\mathcal{A}|a_0 \cap b_1) = \frac{(1-\bar{p})(\underline{p})^m}{(1-\bar{p})(\underline{p})^m + (1-\bar{p})\underline{p}^m},$$

where the expert signal is denoted by s_0. Since the RHS is decreasing to zero in m, $\exists n \in \mathbb{N}^+$ s.t. $P(\mathcal{A}|a_0 \cap a_1) < 1/2$. Then $P(c|v_0, \mathbf{v_n}) < P(a_1 \cap a_2|\mathcal{A}) + P(b_1 \cap a_2|\mathcal{A}) = P(a_2|\mathcal{A}) = \underline{p} < \bar{p}$.

The result for $P(f|v_0, \mathbf{v_n})$ is similar. $\square$

Proof of Proposition 4. The proof is by induction. Assume (S1), (A1), (A2), and $\mathcal{S} = \mathcal{A}$ (the proof for $\mathcal{S} = \mathcal{B}$ is similar). The induction hypothesis clearly holds for $n = 1$. For illustration, consider the general case of $n = 2$ presented in Table 1 of Section IV.

By definition, $x + y = p_2$, and $x > 1 - p_1 - y$ since $p_1, p_2 > 1/2$. If $p_2 > p_1$ (equivalently, $p_1 - x < y$), then $P(f|\mathbf{v_n}) = p_1 - x + 1 - p_1 - y = 1 - p_2 = 1 - \max\{p_1, p_2\}$. If $p_2 < p_1$ (equivalently, $p_1 - x > y$), then $P(f|\mathbf{v_n}) = y + 1 - p_1 - y = 1 - p_1 = 1 - \max\{p_1, p_2\}$. Lastly, if $p_2 = p_1$, then $P(f|\mathbf{v_n}) = 1 - p_1 - y < 1 - p_1 = 1 - \max\{p_1, p_2\}$. Therefore, $P(f|\mathbf{v_n}) \leq 1 - \max\{p_1, p_2\}$, and the induction hypothesis holds for $n = 2$.

Next consider the general case for $n > 2$. Index the n^2 possible signal vectors by $\mathbf{s}_j$ for $j = 1, 2, ..., n^2$. Without loss of generality, order the signal vectors according to whether they create a false or correct inference, with the uninformative cases at the end:

$$P(\mathbf{s}_j|\mathcal{A}) = \begin{cases} f_j \text{ for } j = 1, 2, ..., k \\ e_j \text{ for } j = k+1, ..., n^2/2 \\ c_j \text{ for } j = n^2/2 + 1, ..., n^2/2 + k + 1 \\ e_j \text{ for } j = n^2/2 + k + 2, ..., n^2 \end{cases}$$

for some $k \in \{1, 2, ..., n^2/2\}$ (*Proposition 3* shows $k \geq 1$), where it is understood that if $k = n^2/2$, no event is uninformative. Furthermore, let $\mathbf{s}_j = \bar{\mathbf{s}}_{j+n^2/2}$ for $j = 1, 2, ..., n^2/2$. Define f_{aj}, e_{aj}, and c_{aj} for the respective probabilities of the intersections with a_{n+1}. Likewise, define f_{bj} for the conditional intersection of b_{n+1} and so on. Then

$$\begin{aligned} P(f|\mathbf{v_{n+1}}) = \sum_{j=1}^{k} & \mathbb{1}(f_{aj} < c_{bj+n^2/2}) f_{aj} + \mathbb{1}(f_{bj} < c_{aj+n^2/2}) f_{bj} \\ & + \mathbb{1}(c_{aj+n^2/2} < f_{bj}) c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} < f_{aj}) c_{bj+n^2/2} \\ + \sum_{j=k+1}^{n^2/2} & \mathbb{1}(e_{aj} < e_{bj+n^2/2}) e_{aj} + \mathbb{1}(e_{bj} < e_{ej+n^2/2}) e_{bj} \\ & + \mathbb{1}(e_{aj+n^2/2} < e_{bj}) e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} < e_{aj}) e_{bj+n^2/2}. \end{aligned}$$

Because

$$\mathbb{1}(f_{aj} < c_{bj+n^2/2}) f_{aj} + \mathbb{1}(f_{bj} < c_{aj+n^2/2}) f_{bj} + \mathbb{1}(c_{aj+n^2/2} < f_{bj}) c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} < f_{aj}) c_{bj+n^2/2} \\ \leq f_{bj} + c_{bj+n^2/2},$$

$$\mathbb{1}(e_{aj} < e_{bj+n^2/2}) e_{aj} + \mathbb{1}(e_{bj} < e_{aj+n^2/2}) e_{bj} + \mathbb{1}(e_{aj+n^2/2} < e_{bj}) e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} < e_{aj}) e_{bj+n^2/2} \\ \leq e_{bj} + e_{bj+n^2/2},$$

and

$$1 - p_{n+1} = \sum_{j=1}^{k} (f_{bj} + c_{bj+n^2/2}) + \sum_{j=k+1}^{n^2/2} (e_{bj} + e_{bj+n^2/2})$$

by construction, then $P(f|\mathbf{v_{n+1}}) \leq 1 - p_{n+1}$.

Next, because

$\mathbb{1}(f_{aj} < c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} < c_{aj+n^2/2})f_{bj} + \mathbb{1}(c_{aj+n^2/2} < f_{bj})c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} < f_{aj})c_{bj+n^2/2}$
$\leq f_j$, and

$\mathbb{1}(e_{aj} < e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} < e_{aj+n^2/2})e_{bj} + \mathbb{1}(e_{aj+n^2/2} < e_{bj})e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} < e_{aj})e_{bj+n^2/2}$
$\leq e_j$,

the induction hypothesis implies (via *Lemma 3*):

$$P(f|\mathbf{v_{n+1}}) \leq \sum_{j=1}^{k} f_j + \sum_{j=k+1}^{n^2/2} e_j \leq 1 - p^*.$$

Thus, $P(f|\mathbf{v_{n+1}}) \leq 1 - \max\{p^*, p_{n+1}\}$. □

Lemma 3. Assume (S1), (A1), and (A2). Adopting the terminology and ordering of *Proposition 4*,

$$\sum_{j=1}^{k} f_j + \sum_{j=k+1}^{n^2/2} e_j \leq 1 - p^* \; \forall n \in \mathbb{N}^+.$$

Proof of Lemma 3. Assume (S1), (A1), and (A2). The proof is by induction. The induction hypothesis clearly holds for $n = 1$. Suppose $n > 1$. Adopting the terminology and ordering of *Proposition 4*, we seek to show

$$\begin{aligned}
\Gamma \equiv \sum_{j=1}^{k} & \mathbb{1}(f_{aj} < c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} < c_{aj+n^2/2})f_{bj} \\
& + \mathbb{1}(c_{aj+n^2/2} < f_{bj})c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} < f_{aj})c_{bj+n^2/2} \\
& \quad + \mathbb{1}(f_{aj} = c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} = c_{aj+n^2/2})f_{bj} \\
& + \sum_{j=k+1}^{n^2/2} \mathbb{1}(e_{aj} < e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} < e_{ej+n^2/2})e_{bj} \\
& + \mathbb{1}(e_{aj+n^2/2} < e_{bj})e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} < e_{aj})e_{bj+n^2/2} \\
& \quad + \mathbb{1}(e_{aj} = e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} = e_{ej+n^2/2})e_{bj} \\
& \leq 1 - \max\{p^*, p_{n+1}\}.
\end{aligned}$$

Because

$$\mathbb{1}(f_{aj} < c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} < c_{aj+n^2/2})f_{bj} + \mathbb{1}(c_{aj+n^2/2} < f_{bj})c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} < f_{aj})c_{bj+n^2/2}$$
$$+ \mathbb{1}(f_{aj} = c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} = c_{aj+n^2/2})f_{bj} \leq f_{bj} + c_{bj+n^2/2},$$

$$\mathbb{1}(e_{aj} < e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} < c_{aj+n^2/2})e_{bj} + \mathbb{1}(e_{aj+n^2/2} < f_{bj})e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} < e_{aj})e_{bj+n^2/2}$$
$$+ \mathbb{1}(e_{aj} = e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} = e_{aj+n^2/2})e_{bj} \leq e_{bj} + e_{bj+n^2/2},$$

and

$$1 - p_{n+1} = \sum_{j=1}^{k}(f_{bj} + c_{bj+n^2/2}) + \sum_{j=k+1}^{n^2/2}(e_{bj} + e_{bj+n^2/2})$$

by construction, then $\Gamma \leq 1 - p_{n+1}$.

Next, because

$$\mathbb{1}(f_{aj} < c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} < c_{aj+n^2/2})f_{bj} + \mathbb{1}(c_{aj+n^2/2} < f_{bj})c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} < f_{aj})c_{bj+n^2/2}$$
$$+ \mathbb{1}(f_{aj} = c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} = c_{aj+n^2/2})f_{bj} \leq f_j,$$

and

$$\mathbb{1}(e_{aj} < e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} < c_{aj+n^2/2})e_{bj} + \mathbb{1}(e_{aj+n^2/2} < f_{bj})e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} < e_{aj})e_{bj+n^2/2}$$
$$+ \mathbb{1}(e_{aj} = e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} = e_{aj+n^2/2})e_{bj} \leq e_j,$$

the induction hypothesis implies

$$\Gamma \leq \sum_{j=1}^{k} f_j + \sum_{j=k+1}^{n^2/2} e_j \leq 1 - p^*.$$

Thus, $\Gamma \leq 1 - \max\{p^*, p_{n+1}\}$. □

Proof of Proposition 5. Assume (S1), (A1), and (A2). Following the same approach and terminology of *Proposition 4*, the induction hypothesis clearly holds for $n = 1$. For $n + 1$, we have

$$\begin{aligned} P(e|\mathbf{v_{n+1}}) = \sum_{j=1}^{k} & \mathbb{1}(f_{aj} = c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} = c_{aj+n^2/2})f_{bj} \\ & + \mathbb{1}(c_{aj+n^2/2} = f_{bj})c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} = f_{aj})c_{bj+n^2/2} \\ + \sum_{j=k+1}^{n^2/2} & \mathbb{1}(e_{aj} = e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} = e_{ej+n^2/2})e_{bj} \\ & + \mathbb{1}(e_{aj+n^2/2} = e_{bj})e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} = e_{aj})e_{bj+n^2/2}. \end{aligned}$$

Because

$$\mathbb{1}(f_{aj} = c_{bj+n^2/2})f_{aj} + \mathbb{1}(f_{bj} = c_{aj+n^2/2})f_{bj} + \mathbb{1}(c_{aj+n^2/2} = f_{bj})c_{aj+n^2/2} + \mathbb{1}(c_{bj+n^2/2} = f_{aj})c_{bj+n^2/2}$$
$$\leq 2f_{bj} + 2c_{bj+n^2/2},$$

$\mathbb{1}(e_{aj} = e_{bj+n^2/2})e_{aj} + \mathbb{1}(e_{bj} = e_{aj+n^2/2})e_{bj} + \mathbb{1}(e_{aj+n^2/2} = e_{bj})e_{aj+n^2/2} + \mathbb{1}(e_{bj+n^2/2} = e_{aj})e_{bj+n^2/2}$
$\leq 2e_{bj} + 2e_{bj+n^2/2}$,

and

$$1 - p_{n+1} = \sum_{j=1}^{k}(f_{bj} + c_{bj+n^2/2}) + \sum_{j=k+1}^{n^2/2}(e_{bj} + e_{bj+n^2/2})$$

by construction, then $P(e|\mathbf{v_{n+1}}) \leq 2(1-p_{n+1})$. An exact analogy to *Lemma 3* then establishes $P(e|\mathbf{v_{n+1}}) \leq 2(1-p^*)$ (proof omitted). Thus, $P(e|\mathbf{v_{n+1}}) \leq 2(1-\max\{p^*, p_{n+1}\})$ □

Proof of Lemma 4. Since $F(p,n)$ is differentiable, proving concavity is equivalent to showing its second derivative,

$$\frac{\partial^2 F(p,n)}{\partial p^2} = \sum_{k=\frac{(n+1)}{2}}^{n} \binom{n}{k} \Big[k(k-1)p^{k-2}(1-p)^{n-k} - 2k(n-k)p^{k-1}(1-p)^{n-k-1} + (n-k)(n-k-1)p^k(1-p)^{n-k-2}\Big],$$

is strictly negative over $p \in (1/2, 1)$. First, reversing the order of the summation,

$$F(p,n) = \binom{n}{n}p^n + \binom{n}{n-1}p^{n-1}(1-p) + \binom{n}{n-2}p^{n-2}(1-p)^2 + ... + \binom{n}{\frac{n+1}{2}}p^{n-(n+1)/2}(1-p)^{(n-1)/2}.$$

Then

$$\frac{\partial F(p,n)}{\partial p} = np^{n-1} + \frac{n!}{(n-1)!}\Big[(n-1)p^{n-2}(1-p) - p^{n-1}\Big] + \frac{n!}{2!(n-2)!}\Big[(n-2)p^{n-3}(1-p)^2 - 2p^{n-2}(1-p)\Big] + ... + \frac{n!}{\left(\frac{n+1}{2}\right)!\left(\frac{n-1}{2}\right)!}\Big[((n+1)/2)p^{(n+1)/2-1}(1-p)^{(n-1)/2} - ((n-1)/2)p^{n-1}(1-p)^{(n-1)/2-1}\Big]$$

is a telescoping series, leaving

$$\frac{\partial F(p,n)}{\partial p} = \frac{n!}{\left(\frac{n+1}{2}\right)!\left(\frac{n-1}{2}\right)!}\Big[((n+1)/2)p^{(n+1)/2-1}(1-p)^{(n-1)/2}\Big] = \frac{n!}{\left(\left(\frac{n-1}{2}\right)!\right)^2}\big[p(1-p)\big]^{(n-1)/2}.$$

Therefore,

$$\frac{\partial^2 F(p,n)}{\partial p^2} = \frac{n!}{\left(\left(\frac{n-1}{2}\right)!\right)^2}\Big[((n-1)/2)(p(1-p))^{(n-1)/2-1}(1-2p)\Big],$$

and

$$\frac{\partial^2 F(p,n)}{\partial p^2} < 0 \iff (1-2p) < 0 \iff 1/2 < p. \quad \square$$

Proof of Proposition 6. By the proof of *Lemma 3*,

$$\frac{\partial F(p,n)}{\partial p} = \frac{n!}{\left(\left(\frac{n-1}{2}\right)!\right)^2}\left[p(1-p)\right]^{(n-1)/2}.$$

Then

$$\frac{\partial^2 F(p,n+2)}{\Delta n \partial p} = \frac{(n+2)!}{\left(\left(\frac{n+1}{2}\right)!\right)^2}\left[p(1-p)\right]^{(n+1)/2} - \frac{n!}{\left(\left(\frac{n-1}{2}\right)!\right)^2}\left[p(1-p)\right]^{(n-1)/2},$$

which, after some algebra, is greater than zero if and only if $4p(1-p) > (n+1)/(n+2)$. Since $4p(1-p) = 1$ at its maximizer $p = 1/2$, solving the corresponding quadratic equation yields

$$\frac{\partial^2 F(p,n+2)}{\Delta n \partial p} \begin{cases} \geq 0 \text{ if } p \in [1/2, 1/2 + \sqrt{1-(n+1)/(n+2)}] \\ \leq 0 \text{ if } p \in [1/2 + \sqrt{1-(n+1)/(n+2)}, 1]. \end{cases}$$

Since $EU(p,n)$ is an affine function of $F(p,n)$, the optimal group size is single-peaked. $\square$

Proof of Lemma 5. Writing the optimal investment as a function of the group size, the first-order condition implies

$$\frac{\partial F(p^*(n),n)}{\partial p} uN - n\phi'(p^*(n)) = 0$$

for any n. Dividing the equality for an incremental group member yields

$$\frac{\frac{\partial F(p^*(n+2),n+2)}{\partial p}}{\frac{\partial F(p^*(n),n)}{\partial p}} = \frac{(n+2)\phi'(p^*(n+2))}{n\phi'(p^*(n))}. \tag{L5}$$

Now suppose $p^*(n+2) \geq p^*(n)$. Then the RHS of (L5) is strictly greater than one. By *Proposition 1*, $F(p,n)$ is increasing and approaching one as n grows. Furthermore, the difference $F(p,n+2) - F(p,n)$ is decreasing in n. By the proof of *Lemma 3*, $F(p,n)$ is concave and approaches one as p grows. Thus, the LHS of (L5) must be strictly less than one if $p^*(n+2) \geq p^*(n)$, a contradiction. $\square$