Real People Aggregating Signals:

An Experiment and a Short Story

Michele Piccione

London School of Economics

Ariel Rubinstein

School of Economics, Tel Aviv University
and Department of Economics, New York University

July 30, 2022

ABSTRACT: A decision maker observes multiple signals about an event. He has information about the frequency of the event given each individual signal and wishes to update his beliefs about the event. We examine the problem experimentally and identify some of the commonly used procedures for signal aggregation. These procedures are for the most part inconsistent with Bayesian updating. We apply some of these procedures to a well-known panel game that has previously been studied under the standard Bayesian assumptions and find that the properties of the equilibria differ significantly.

KEYWORDS: signal aggregation, experimental economics, voting game
JEL classification: C9, D70, D80

We are grateful to Ayala Arad whose comment on the negative effects of a positive yet weaker additional signal triggered this project and to Aron Tobias for his insightful and generous comments. We are indebted to the Centre for Behavioural and Experimental Social Science at The University of East Anglia, and especially to Theodore Turocy, and to the Center for Experimental Social Science at New York University, and especially to Anwar Ruff and Shayne Trotman, for their generosity.

1. Introduction

We begin by asking you to answer the following question:

Q1: A very small proportion of the newborns in a certain country have a specific genetic trait.
Two screening tests, A and B, have been introduced for all newborns to identify this trait. However, the tests are not precise. A study has found that:
70% of the newborns who are found to be positive according to test A have the genetic trait (and conversely 30% do not).
20% of the newborns who are found to be positive according to test B have the genetic trait (and conversely 80% do not).
The study has also found that when a newborn has the genetic trait, a positive result in one test does not affect the likelihood of a positive result in the other. Likewise, when a newborn does not have the genetic trait, a positive result in one test does not affect the likelihood of a positive result in the other.

Suppose that a newborn is found to be positive according to both tests. What is your estimate of the likelihood (in %) that this newborn has the genetic trait?

To answer the question you had to process distinct signals, a common operation in everyday life and also in many economic models. The usual story in economic theory is that an agent starts out with a prior belief about relevant unknowns (states of the world), understands the relationship between these unknowns and the signals or events that he has observed, and finally uses Bayes' rule to update his beliefs.

Introspection - supported by a large psychological literature - casts doubt on the hypothesis that Bayesian updating procedures resemble the actual procedures used in real life, even when the information needed to form Bayesian posteriors is clearly specified. In particular, a great deal of attention has been given to the "base-rate fallacy", identified in Kahneman and Tversky (1973) and elucidated in Bar-Hillel (1980).

In the experimental part of the paper, we report the results of a series of experiments in which subjects answered questions like Q1. The vast majority of respondents aggregated the signals using a small number of procedures based on simple formulae that have no similarity to Bayesian updating. Our guess is that your answer to Q1 was one of the following: 14, 20, 45, 70 or 76. Only the last answer is qualitatively consistent with Bayesian updating: since each positive test result strictly increases the likelihood that the newborn has the trait above the base rate, the answer should be strictly above 70.

Note an important feature of the information presented in Q1. In standard economic models, the information structure is usually described by specifying a prior distribution of the states of the world and the distribution of the signals given each state. In contrast, question Q1 specifies the distribution of the states of the world given an outcome of each signal, namely the likelihood of the genetic trait given a positive result in either test. Framing the information structure in this manner is representative of situations where decision makers have access to reliable statistics about the predictive success of individual signals but not of joint ones.

If a decision maker observes only one signal, then his inference is trivial: the likelihood of the trait is given in the description of the signal. Neither the base-rate information nor the distribution of the signal given the states of the world is of any use. However, in Q1 (as well as in other questions discussed below), there are two signals and Bayesian updating is not entirely trivial. The standard Bayesian analysis, presented in Section 2, yields some surprising conclusions. In particular, when the trait is rare, two positive results in conditionally independent tests indicate a very high probability that the individual carries the trait, even when each individual test is unreliable in that a large proportion of those who test positive on one test do not have the trait. This conclusion could have some significance for the legal sphere: two claims about an unlikely event, each unpersuasive on its own, can be very persuasive evidence jointly. For a discussion of aggregation in law see Porat and Posner (2012).

At this point we wish to clarify a methodological point about the interpretation of experimental data. Our objective is not to find the"real-life" distribution of updating procedures in inference problems such as Q1. Our experimental results, and in our opinion any results in experimental economics, can be used only to confirm intuitions. In this paper, it is that many individuals aggregate signals using a small number of simple and identifiable procedures that are incompatible with the Bayesian approach.

The second part of the paper provides a theoretical analysis of non-Bayesian procedures of signal aggregation in a conventional model. We discuss a special case of the model analysed in Duggan and Martinelli (2001): A panel is to determine whether a defendant is guilty or not. The panel consists of n referees each of whom receives a private signal about the guilt of the defendant and then votes whether or not to convict him. A conviction requires unanimity. According to the standard analysis, a referee votes to convict whenever he concludes that, conditional on his signal and on being pivotal (namely, all the other referees voted to convict), the Bayesian posterior probability that the defendant is guilty is sufficiently large. The game has an informative equilibrium where each referee votes to convict if his signal is above some cutoff. One of the conclusions reached is that under some assumptions on the distributions of signals, which we also adopt, when the number of referees is large the voting mechanism yields the correct decision with a probability that approaches 1.

This model serves as our point of departure to examine the extent to which its conclusions hold when the referees are not Bayesian but rather use one of the procedures observed in our experiments. More precisely, we assume that each referee aggregates two signals: his private signal and the event that he is pivotal. Each referee has access to information about the frequency of guilt given either signal separately, applies a signal-aggregation procedure in order to form a belief about the defendant's guilt, and votes to convict if his belief is above some threshold.

Replacing the Bayesian approach with more realistic methods of signal aggregation can yield dramatically different results. We focus on two procedures that were observed in our experiments: the Avg procedure which averages the probabilities of the event given each signal, and the Max procedure which selects the maximum probability. We will show that in equilibrium, as the number of referees increases:

(i) If all referees use the Avg procedure, then the defendant is almost never convicted.

(ii) If all referees use the Max procedure, then the probability of conviction of a guilty defendant converges to 1 and that of convicting an innocent defendant converges to a positive probability.

(iii) The level of welfare converges to the same level in both procedures.

We also consider a case in which the procedures used by the referees are heterogeneous and show that if a fixed, positive proportion of the referees, no matter how small, uses the Avg procedure and the remaining ones are Bayesian then in equilibrium the defendant is almost never convicted as the number of referees becomes large.

We chose this model because it is simple and well known. Many other applications could have been considered. Non-Bayesian procedures of signal aggregation can be directly applied to extensive games with incomplete information where the players move once and sequentially. The two signals are a player's type and the occurrence of an information set at which he makes a decision. For example in "take it or leave it" bargaining between a seller (the proposer) and a buyer (the responder) in which both have private information about quality of the good, the buyer aggregates two signals: his own signal and the one inferred from the seller's offer.

Therefore, the lessons from our exercise extend to other models as well. In mechanism design, auctions, bargaining, and pricing with rational expectations, agents aggregate multiple signals that are either observed directly or inferred from equilibrium as in the panel model. Any claim for the relevance of these models rests on the extent to which their results remain valid when agents use realistic procedures for signal aggregation. Our results cast some doubts.

2. Bayesian analysis of signal aggregation

Suppose that we are interested in ascertaining whether a person has a particular genetic trait and multiple conditionally independent screening tests, denoted by $1,\ldots,K$, are available. Also suppose that eventually it will become known who has the trait and that a proportion $\phi_k \in [s,1)$ of the individuals who tested positive according to test k indeed have it, where $s \in (0,1)$ is the frequency of the trait in the population. Denote by $\pi(K)$ the probability that an individual who is found positive in all K tests indeed has the trait. Define p_k and n_k to be the probabilities of a positive result in test k for an individual with and without the trait, respectively. By standard Bayesian updating, $\frac{\phi_k}{1-\phi_k} = \frac{sp_k}{(1-s)n_k}$ and thus:

$$\frac{\pi(K)}{1-\pi(K)} = \frac{s\prod_k p_k}{(1-s)\prod_k n_k} = \left(\frac{1-s}{s}\right)^{K-1} \prod_k \frac{\phi_k}{1-\phi_k}.$$

Therefore, if there are at least two signals ($K > 1$), $\pi(K)$ depends on s. Three qualitative conclusions based on the above follow:

(i) A test j for which $s = \phi_j$ can be ignored.

(ii) An additional positive result increases the probability that the individual has the trait.

(iii) Keeping all ϕ_k constant, $\pi(K)$ converges to 1 as s converges to 0 for any $K \geq 2$.

Thus, when s is small, a positive result on two conditionally independent tests can significantly increase the probability of the individual having the trait. The conclusion is striking: multiple tests, each not particularly persuasive on its own, can be very persuasive together.

While the above conclusions are fairly intuitive, a different feature of Bayesian inference is less so. Assume that $p_k/n_k = \rho > 1$ for all k. Given a sequence of positive results, the (positive) *marginal* contribution to the posterior probability of an additional positive result is not monotonic. To see this, note that $\frac{\pi(K)}{1-\pi(K)} = \frac{s\rho^K}{1-s}$. Thus, $\pi'(K) = \ln(\rho)\pi(K)(1-\pi(K))$. This derivative is increasing as long as $\pi(K) < \frac{1}{2}$ and decreasing otherwise, and the point of inflection occurs at a value of K in the vicinity of $\frac{\ln((1-s)/s)}{\ln(\rho)}$. As an illustration, the graph of $\pi(K)$ for $s = 0.001$ and $\rho = 5$ demonstrates that most of the gain in posterior probability occurs between $K = 3$ and $K = 6$:

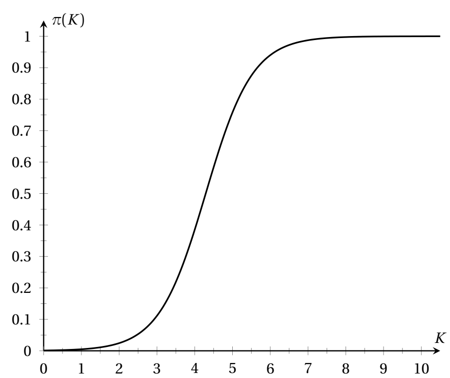

Figure 1: Posterior probability conditional on all tests coming back positive as a function of the number of tests for $s = 0.001$ and $\rho = 5$.

3. Common formulae for signal aggregation

The experimental results point to four formulae used by subjects for aggregating two positive signals:

M^c	$1-(1-\phi_1)(1-\phi_2)$
Avg	$(\phi_1+\phi_2)/2$
Max	$max\{\phi_1,\phi_2\}$
M	$\phi_1\phi_2$

The formulae depend only on ϕ_1 and ϕ_2, thus suggesting base-rate neglect. However, this is not a necessary conclusion since the subjects' selection of a formula may depend on the prior distribution. For instance, if $s = \phi_1 < \phi_2$ then the Max formula coincides with Bayesian updating and is consistent with the understanding that the first positive result conveys no useful information. In contrast, if $s < \phi_1 < \phi_2$, Bayesian updating yields an answer strictly above $max\{\phi_1, \phi_2\}$. In this case, M^c is the only formula in the table that is consistent with the understanding that each positive result increases the posterior probability.

The M^c formula, the use of which appears to be rare, is especially interesting. We conjecture that its rationale is as follows: Since for each test k the proportion of those who have a positive result but do not have the trait is $1-\phi_k$, by "conditional independence", the proportion of individuals with two positive results who do not have the trait is $(1-\phi_1)(1-\phi_2)$ and the proportion of those who do is the residual. This logic would be correct if two independent tests were conducted on two different individuals and we wish to know the probability of at least one of them having the trait. In our context, this logic is incorrect and possibly reflects confusion about the assumption of conditional independence. The formula M is an alternative version of the above (incorrect) logic. However, unlike M^c it does not exhibit the understanding that the posterior is above both ϕ_1 and ϕ_2.

4. The experiments

We conducted a series of experiments to identify common approaches to signal aggregation in problems such as question Q1. Each question in the experiments contains:
I. A background story about a characteristic that is present in a population and either qualitative or quantitative information about its frequency ("a very small proportion" or a specific percentage, respectively).
II. Information about two "screening" tests - for each test, the percentage of individuals with the characteristic among those who test positive.
III. A non-technical explanation of conditional independence between the two tests.
IV. A question: What is your estimate of the likelihood that an individual who is found positive according to both tests has the characteristic?

The experimental research developed in two stages. First, we conducted the experiments on the platform arielrubinstein.org/gt, a website designed for carrying out pedagogical experiments in choice theory and game theory. The subjects were mostly current and past students in game theory courses. No monetary incentives were provided other than a few subjects being randomly chosen to receive $40 regardless of their answers. For us, this stage would have been sufficient. We nonetheless decided to carry out additional experiments based on more standard norms in experimental economics even though the benefit in our case was likely to be negligible. The experiments carried out in the first stage are referred to as the pilot.

The final experiments were conducted through the Center for Experimental Social Science at New York University and the Centre for Behavioural and Experimental Social Science at the University of East Anglia. Students registered with the labs were invited to participate in a short online experiment. Each participant was assigned randomly to one of five questions: Q1, Q2, ..., Q5. We do not report the results for each center separately since the differences were minor.

Q1 which is presented in the Introduction conveys the basic flavor of the experiments. Recall that the question concerns a **genetic trait** which affects a **very small proportion** of **newborns**. Two conditionally independent screening tests are available. The likelihood that a newborn has this trait when testing positive is **70%** on one test and **20%** on the other. The subject is asked to estimate the chances of a newborn having the trait if he tests positive on both tests.

In Q2 we changed the probabilities to **80%** and **60%**. In Q3, we changed the underlying story somewhat by replacing newborns with **undergraduates** who are either continuing on to graduate school or not. A positive result on a test is replaced by a student having taken a certain course. The proportion of students who continue to graduate school is 70% for those who took one course and 20% for those who took the other. The results for Q1, Q2 and Q3 are presented in Table 1.

	mean	n (MRT)	M	Avg	Max	$> Max$ (M^c)	*Other*
Q1 genetic, 70-20	51%	93 (118s)	14%	20%	14%	20% (4%)	32%
Q2 genetic, 80-60	68%	91 (107s)	21%	27%	15%	18% (11%)	19%
Q3 students, 70-20	55%	97 (98s)	7%	27%	20%	21% (6%)	25%

Table 1: Results of Q1, Q2 and Q3.

As is evident from the table, a clear majority (about 60%) of the subjects used one of the four formulae: M, Avg, Max or M^c. Around 20% chose an answer strictly above the maximum of ϕ_1 and ϕ_2, which is qualitatively correct. There do not appear to be significant differences in the results among the three questions. The median response time (MRT) was also quite similar in all three, ranging from 98 to 118 seconds.

It is worth comparing the above results with those obtained using the pedagogical site (the pilot). As is evident from Table 2, which only relates to Q1, the results are similar. The differences were: (i) students in the pilot have a *higher* response time (the MRT was in the range of 120-200 seconds for several questions); (ii) a larger proportion of subjects in the pilot (31% vs. 20%) gave a qualitatively correct answer (a number exceeding both ϕ_1 and ϕ_2).

	mean	n (MRT)	M	Avg	Max	$> Max$ (M^c)	*Other*
Labs	51%	93 (118s)	14%	20%	14%	20%(4%)	32%
Pilot	51%	74 (147s)	18%	11%	9%	31%(12%)	31%

Table 2: Q1 - Labs vs. Pilot.

Q4 is used to determine the consistency of the approaches used by the subjects. We asked the subjects to simultaneously assess the likelihood of the trait given two different pairs of tests: Alice tested positive on two tests with accuracies of 70% and 20% respectively and Bob tested positive on two tests with accuracies of 50% and 40% respectively. We observe a considerable degree of consistency. Out of 92 subjects about 60% were consistent in their use of one of four formulae: M (15%), Avg (30%), Max (13%) and M^c (2%). Furthermore, 8% of subjects consistently chose answers above Max that differed from the answer according to M^c.

Q5 is identical to Q1 except that the base rate is 20% (and not "a very small proportion"). As noted earlier, when the base rate is equal to ϕ_1, the correct answer is ϕ_2 since subjects should ignore a positive test result with the same accuracy as the base rate. Nonetheless, all four formulae are still used in this question. Notably, there is no significant change in the use of Max, which gives the correct answer, relative to Q1. About 13% of the subjects gave the base-rate probability (20%) as their answer. Note also that 16% of subjects chose an answer strictly above both ϕ_1 and ϕ_2, although in this question any answer different from 70% is qualitatively incorrect.

	mean	n (MRT)	M	20%	Avg	Max	$> Max$ (M^c)	*Other*
Q5	49%	94 (121s)	10%	13%	9%	17%	16% (3%)	36%

Table 3: Q5 results.

Remark (Bar-Hillel (1980)): A version of Problem 6 in Bar-Hillel (1980) is closest to our questions. However, in her setup information about the signals is presented as in standard economic models, namely by specifying their precision conditional on the true state.[1] In Bar-Hillel (1980) 83% of the subjects chose the average (75%) and thus "were

[1] "Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). 85% of the cabs in the city are Blue, and the remaining 15% are Green. A cab was involved in a hit-and-run accident at night. There were two witnesses to the accident. Both claimed that the errant cab had been Green. The court tested the witnesses' ability to distinguish between Blue and Green cabs under night-time visibility conditions. It found the first witness able to identify the correct color about 80% of the time, confusing it with the other color 20% of the time; the second witness identified each color correctly 70% of the time, and erred about 30% of the time. What do you think are the chances that the errant cab was Green?"

disregarding the base rate". The number of participants was low (29). We repeated the experiment in our pilot with a somewhat larger pool of subjects (93) and, as in Bar-Hillel (1980), without monetary incentives. Our results were different: only 16% of the subjects answered 75% and the other answers were widely dispersed: 8% chose 56%, 8% chose 15% and only 5% gave the correct answer (62%).

5. A panel story

In the remainder of the paper we investigate the effects of using some of the non-Bayes -ian procedures observed in our experiments by means of a panel game analysed conventionally in Duggan and Martinelli (2001).

A panel of n referees is deciding whether a defendant is guilty. The prior probability of him being guilty is s and that of being innocent is $1-s$. Each referee votes either Y (guilty) or N (not guilty) and the defendant is found guilty when all the referees vote Y. The referees vote simultaneously. Prior to voting each referee receives a private signal in the form of a number in the interval $[0,1]$. The signals are identically distributed and conditionally independent across the referees. The cdf of each signal conditional on the defendant being guilty is F and that conditional on being innocent is G. The cdfs F and G have continuous density functions f and g, respectively. We impose the following restrictions throughout:

(i) $f(0)=0$, $g(1)=0$, $f(t)>0$ for all $t>0$ and $g(t)>0$ for all $t<1$.

(ii) $\frac{f(t)}{g(t)}$ is strictly increasing.

Note that (ii) implies that $\frac{F(t)}{G(t)} < \frac{f(t)}{g(t)} < \frac{1-F(t)}{1-G(t)}$ for all $t \in (0,1)$, that $\frac{F(t)}{G(t)}$ and $\frac{1-F(t)}{1-G(t)}$ are strictly increasing, and together with (i), that $\lim_{t\to 1} \frac{1-F(t)}{1-G(t)} = \infty$.

Each referee prefers that the defendant is convicted if he believes that the probability of him being guilty is at least z, a number in $(0,1)$. This is consistent with each referee maximizing a vNM utility that is equal to 1 if the correct decision is made when the defendant is guilty, to $\lambda = \frac{z}{1-z}$ if the correct decision is made when the defendant is innocent and to 0 if the incorrect decision is made. Therefore, the natural welfare function when all n referees use a common cutoff α, namely each referee votes Y if and only if his observed signal is at least α, is:

$$W^n(\alpha) = s(1-F(\alpha))^n + (1-s)\lambda(1-(1-G(\alpha))^n).$$

We focus on the case where $z \geq 1/2 \geq s$. That is, the ex-ante belief that the defendant is guilty is not higher than the belief that he is innocent and the standards for conviction are higher than the standards for acquittal.

5.1 Signals and beliefs

In our experiments, the signals are exogenous events that are communicated to the agents. In a strategic environment with incomplete information, the strategies of the players generate events that can be interpreted as "signals" and they can be aggregated with other exogenous signals to formulate relevant posterior beliefs.

In the standard approach to binary voting models a voter's best reply depends on his own information and the event in which his vote is pivotal. Following this approach we assume that a referee forms his beliefs by aggregating two "signals": (i) his own private signal and (ii) the occurrence of the event that he is pivotal. Note that in this model the signal about a referee's vote being pivotal is independent of the referee's own vote. This feature allows the introduction of strategic considerations into non-Bayesian procedures in an uncomplicated way.

We refer to referee i's belief that the defendant is guilty given his private signal t and conditional on him being pivotal given a strategy profile S_{-i} of the other players as his *C-belief* and denote it by $\mu_i(t, S_{-i})$. If all the players use the same cutoff σ, namely each votes Y if and only if his signal is above σ, then we denote the C-belief by $\mu_i(t,\sigma)$. Naturally, C-beliefs depend on the procedures - whether Bayesian or non-Bayesian - used by the referees to update their beliefs.

5.2 Equilibrium

To summarize, the model is the tuple $\langle n, s, z, F, G, (\mu_i)\rangle$ where n is the number of referees, s is the common prior probability that the defendant is guilty, z is the common minimal belief as to the guilt of the defendant for which a conviction is optimal, F and G are the distributions of each private signal given that the defendant is respectively guilty or not guilty, and μ_i is referee i's C-belief.

We are left to define the equilibrium concept. A *(symmetric) equilibrium* is a cutoff $\sigma^* < 1$ such that for every referee i we have $\mu_i(t,\sigma^*) \leq z$ for all $t < \sigma^*$ and $\mu_i(t,\sigma^*) \geq z$ for all $t > \sigma^*$. That is, in equilibrium whenever a referee votes N he believes that the probability of the defendant being guilty is sufficiently low (weakly below z) and whenever he votes Y he believes that it is sufficiently high (weakly above z). We require that $\sigma^* < 1$ to exclude the case in which all players vote N regardless of their signals and no player is ever pivotal.

6. Bayesian equilibria

We first review the model under the standard approach where for each i the C-beliefs are defined by Bayesian updating:

$$\mu_i(t,\sigma) = \frac{sf(t)(1-F(\sigma))^{n-1}}{sf(t)(1-F(\sigma))^{n-1} + (1-s)g(t)(1-G(\sigma))^{n-1}}.$$

Since μ_i is strictly increasing in t for a fixed $\sigma < 1$, σ^* is an equilibrium if and only if it satisfies:

$$\mu_i(\sigma^*,\sigma^*) = z.$$

Clearly, the equilibrium is unique.[2] For later use, we rearrange the equilibrium condition. Let $r^k(\theta)$ be the probability that the defendant is guilty conditional on k referees with a common cutoff θ voting Y, that is,

$$r^k(\theta) = \frac{s(1-F(\theta))^k}{s(1-F(\theta))^k + (1-s)(1-G(\theta))^k},$$

and we set $r^k(1) = 1$, which is the limit of $r^k(\theta)$ as $\theta \to 1$. The function r^k is strictly increasing and $r^k(\theta) < r^{k+1}(\theta)$ for any $k \geq 1$ and $0 < \theta < 1$. Then, σ^* is an equilibrium if and only if it satisfies:

$$r^{n-1}(\sigma^*) = \frac{zg(\sigma^*)}{zg(\sigma^*) + (1-z)f(\sigma^*)}.$$

[2]Given our assumptions we have that $\mu_i(0,0) = 0$, the function $\mu_i(t,t)$ is increasing, continuous and converges to 1 as $t \to 1$. Therefore, there is a unique α sastifying the equation $\mu_i(\alpha,\alpha) = z$.

It follows from Duggan and Martinelli (2001) that the equilibrium condition also characterizes welfare maximization with a common cutoff. As the number of referees grows, the optimal cutoff decreases and converges to 0. Furthermore, the probability of an incorrect decision given the optimal cutoff converges to 0. That is, the equilibrium level of welfare in the standard Bayesian approach converges to “the first-best”, namely $s+(1-s)\lambda$.

7. Non-Bayesian games: aggregating two signals

We now consider the case where all or some of the players are not Bayesian and use one of the procedures for signal aggregation observed in the experiments. All these procedures use a formula that is a function of two numbers: the probability of the relevant state conditional on each signal. In our case, the two numbers are:
(i) the probability $p(t)$ of the defendant being guilty given the referee's own signal t, that is:

$$p(t)=\frac{sf(t)}{sf(t)+(1-s)g(t)}$$

and (ii) the probability $r^{n-1}(\theta)$ that the defendant is guilty given that the referee is pivotal and all other referees use the cutoff θ.
Note that the function p is strictly increasing and satisfies $p(t)<r^k(t)$ for any $t\in(0,1)$ and $k\geq 1$.

As mentioned above, in the standard analysis of the panel game the equilibrium cutoff maximizes welfare and when the number of referees is large the correct outcome is obtained almost with certainty. Our results show that these conclusions do not carry over to some of the non-Bayesian procedures observed in our experiments. In particular:
(1) If all referees use the Avg procedure, then the equilibrium cutoff is higher than the Bayesian one and when the number of referees is large almost all defendants are acquitted.
(2) If all referees use the Max procedure, then the equilibrium cutoff will typically be sub-optimal. When the number of referees is large, the equilibrium cutoff is below the Bayesian one and all guilty defendants are convicted as well as some proportion of the

innocent ones. Remarkably, the level of welfare converges to that achieved when the referees never convict.
(3) If some proportion of the referees - no matter how small - uses the Avg procedure and the remaining referees are Bayesian then, when the number of referees goes to infinity, it will be almost certain that the defendant is acquitted.

While a "large" Bayesian panel approximates the first best, all other types of "large" panels we investigate approximately achieve the same level of welfare as a panel that never convicts, even though the members of the panel use different equilibrium strategies.

7.1 The Avg game

Assume that all referees use the Avg procedure, that is referee i's C-belief is:

$$\mu_i(t,\sigma) = \frac{1}{2}(p(t) + r^{n-1}(\sigma)).$$

Since C-beliefs are monotonic in t equilibria are characterized by the solutions of the equation $\frac{1}{2}(p(\beta) + r^{n-1}(\beta)) = z$. The following claim states that the equilibrium cutoff is above the Bayesian cutoff and as the number of referees grows the probability that any defendant is convicted declines to zero.

Claim A *In the Avg game:*
(i) There is a unique equilibrium denoted by β^*.
(ii) $\beta^* \geq \sigma^*$ with strict inequality unless $z = s = \frac{1}{2}$.
(iii) Denote by $\beta^*(n)$ the equilibrium in the game with n referees. If $z > \frac{1}{2}$ then as $n \to \infty$ (a) $\beta^*(n) \to \bar{\beta}$ where $p(\bar{\beta}) + 1 = 2z$ and (b) the level of welfare converges to $\lambda(1-s)$ (the defendant is almost never convicted).

Proof. (i) The function $\frac{1}{2}(p(\beta) + r^{n-1}(\beta))$ is increasing in β and ranges from $\frac{s}{2}$ to 1 and hence there is a unique β^* at which it equals z. Clearly this is the unique equilibrium.

(ii) Since $z \geq s$

$$p(\sigma^*)+r^{n-1}(\sigma^*)=\frac{sf(\sigma^*)}{sf(\sigma^*)+(1-s)g(\sigma^*)}+\frac{zg(\sigma^*)}{zg(\sigma^*)+(1-z)f(\sigma^*)}\leq$$

$$\frac{zf(\sigma^*)}{zf(\sigma^*)+(1-z)g(\sigma^*)}+\frac{zg(\sigma^*)}{zg(\sigma^*)+(1-z)f(\sigma^*)}\leq 2z.$$

The first inequality holds strictly unless $s=z=\frac{1}{2}$ given that $z\geq\frac{1}{2}\geq s$. The second inequality follows from $\frac{x}{zx+(1-z)y}+\frac{y}{zy+(1-z)x}\leq 2$ which holds with equality only for $z=\frac{1}{2}$ or $x=y$ (or $z=1$). Therefore one of the inequalities will hold strictly unless $s=z=\frac{1}{2}$. By the monotonicity of p and r^{n-1}, it follows that $\beta^*\geq\sigma^*$ with strict inequality unless $z=s=\frac{1}{2}$.

(iii) Since the function $\frac{1}{2}(p(\beta)+r^{n-1}(\beta))$ is increasing in β and n, the sequence $\beta^*(n)$ is decreasing. Since $p(\beta^*(n))\geq 2z-1>0$, $\beta^*(n)$ is bounded away from 0 and the equilibrium probability that the defendant will be found guilty converges to 0 as $n\to\infty$. Therefore, the sequence $\beta^*(n)$ converges to the unique solution of $p(\beta)+1=2z$. □

7.2 The Max game

Suppose that all referees use the Max procedure, that is, for all i:

$$\mu_i(t,\theta)=max\{p(t),r^{n-1}(\theta)\}.$$

The following claim states that the equilibrium cutoff is always below the Avg cutoff and also below the Bayesian one when n is large. As the number of referees grows, the equilibrium level of welfare approaches the equilibrium level of welfare in the Avg game, namely the level of welfare in the case of a panel that never convicts, despite the fact that convictions occur with probability larger than s.

Claim B *In the Max game:*
(i) There exists a unique equilibrium γ^* satisfying $r^{n-1}(\gamma^*)=z$.
(ii) $\gamma^*\leq\beta^*$.
(iii) Let $\gamma^*(n)$ be the equilibrium in the game with n referees. There exists a sequence $\eta(n)$ that converges to ∞ such that for any n, $\sigma^*(n)>\gamma^*(n)$ if and only if $\lambda<\eta(n)$.

(iv) As n increases, the probability of convicting a guilty defendant converges to 1 *and that of convicting an innocent one converges to* $\frac{s}{(1-s)\lambda}$. *The level of welfare converges to* $\lambda(1-s)$.

Proof. (i) Let γ^* be the unique solution of $r^{n-1}(\gamma)=z$. Then γ^* is an equilibrium: if a referee receives a signal below γ^* his C-belief is z and thus he is indifferent between voting Y and voting N; if he receives a signal above γ^*, then his C-belief is at least z and thus voting Y is optimal.

There is no other equilibrium:

(a) A common cutoff $\underline{\gamma}<\gamma^*$ is not an equilibrium since a referee with a signal $t\in(\underline{\gamma},\gamma^*)$ has a C-belief equal to $max\{p(t),\ r^{n-1}(\underline{\gamma})\}$ which is less than z since $p(t)<p(\gamma^*)<r^{n-1}(\gamma^*)=z$ and $r^{n-1}(\underline{\gamma})<r^{n-1}(\gamma^*)=z$. Such a referee prefers to vote N.

(b) A common cutoff $\bar{\gamma}>\gamma^*$ is not an equilibrium since in that case any referee's C-belief is at least $r^{n-1}(\bar{\gamma})>z$ and thus voting N is not optimal given any signal.

(ii) The assertion follows from the Avg equilibrium condition $\frac{1}{2}(p(\beta^*)+r^{n-1}(\beta^*))=z$ and the fact that $r^{n-1}(t)>p(t)$ for all $t\in(0,1)$.

(iii) $\gamma^*(n)$ is the solution to $\frac{r^{n-1}(t)}{1-r^{n-1}(t)}=\lambda$ and $\sigma^*(n)$ is the solution to $\frac{r^{n-1}(t)}{1-r^{n-1}(t)}\frac{f(t)}{g(t)}=\lambda$. The two LHS functions are increasing and have the same value only at the point $\bar{t}$ such that $f(\bar{t})=g(\bar{t})$. Define $\eta(n)=\frac{r^{n-1}(\bar{t})}{1-r^{n-1}(\bar{t})}$. The sequence converges to infinity since $F(\bar{t})<G(\bar{t})$. Then, $\lambda<\eta(n)$ iff $\gamma(n)<\bar{t}$ iff $\frac{f(\gamma(n))}{g(\gamma(n))}<1$ iff $\sigma^*(n)>\gamma^*(n)$.

(iv) Since $\gamma^*(n)<\sigma^*(n)$ for large n, the probability that a guilty defendant is found guilty goes to 1. The ratio between guilty and innocent defendants that are convicted is $\frac{s(1-F(\gamma^*(n)))^n}{(1-s)(1-G(\gamma^*(n)))^n}=\frac{r^{n-1}(\gamma^*(n))}{1-r^{n-1}(\gamma^*(n))}\frac{1-F(\gamma^*(n))}{1-G(\gamma^*(n))}$ which converges to λ by the equilibrium condition. Therefore, the probability of an innocent defendant being found guilty converges to $\frac{s(1-z)}{(1-s)z}$ and the level of welfare converges to $s+(1-s)\lambda(1-\frac{s}{1-s}\frac{1-z}{z})=\lambda(1-s)$. □

Remark (the *Min* game): By an argument similar to that in Claim B the only equilibrium for the game in which all referees follow the Min procedure is δ^* satisfying $p(\delta^*)=z$. Since the equilibrium is independent of n and $\delta^*>0$, as the number of referees grows, the probability of conviction goes to zero and the level of welfare converges to $(1-s)\lambda$, as in the Avg and Max games.

7.3 Mixed Bayesian and Avg game

Suppose that $n\kappa$ players are Bayesian and the remaining referees are *Avg*. We assume that at least two referees are Avg. An extension of the equilibrium definition specifies that $\alpha^* < 1$ and $\beta^* < 1$ where α^* is the common cutoff of the Bayesian players and β^* is the common cutoff of the Avg players.

Claim C *Suppose that* $z > \frac{1}{2} > s$.[3]
(i) An equilibrium exists.
(ii) Any sequence of equilibria $(\alpha^*(n), \beta^*(n))$ *converges to* $(0, \bar{\beta})$ *where* $p(\bar{\beta}) = 2z - 1$ *and the equilibrium probability of conviction converges to zero as* $n \to \infty$.

Proof. (i) Define $\psi(t) = \frac{1-G(t)}{1-F(t)}$. An equilibrium $(\alpha^*, \beta^*) \in (0,1) \times (0,1)$ is characterized as being a solution to the two equations:

$$\frac{1}{1 + \frac{1-s}{s}\frac{g(\alpha)}{f(\alpha)}(\psi(\alpha))^{\kappa n-1}(\psi(\beta))^{n-\kappa n}} = z$$

$$\frac{1}{2}[p(\beta) + \frac{1}{1 + \frac{1-s}{s}(\psi(\alpha))^{\kappa n}(\psi(\beta))^{n-\kappa n-1}}] = z$$

A solution in $(0,1) \times (0,1)$ exists.[4]

(ii) Let $(\alpha^*(n), \beta^*(n))$ be a sequence of equilibria. The sequence $\alpha^*(n)$ converges to 0. If not there would be an $\epsilon > 0$ and a subsequence that is above ϵ. Since $\psi(\epsilon) \in (0,1)$ and ψ is increasing, the LHS of the first equation along the subsequence would converge to $1 > z$.

[3]When $s = z = \frac{1}{2}$ there exists an equilibrium which is identical to the one in which all n players are Bayesian (or Avg). Let τ^* be the solution of the equation $\frac{g(\tau)}{f(\tau)}(\psi(\tau))^{n-1} = 1$. Then, τ^* also solves the equation $p(\tau) + 1/(1 + (\psi(\tau))^{n-1}) = 1$. Therefore, $\alpha^* = \beta^* = \tau^*$ is an equilibrium. Obviously, τ^* is also an equilibrium when all referees are either Bayesian or Avg.

[4]For every $\beta < 1$ there is a unique $\alpha \in [0,1)$ that solves the first equation since the LHS of the equation converges to 0 as $\alpha \to 0$ and to 1 as $\alpha \to 1$. Denote this solution by $a(\beta)$. For every $\alpha \in [0,1]$ there is a unique $\beta \in (0,1)$ which solves the second equation since the LHS converges to 1 as $\beta \to 1$ and is equal to $\frac{1}{2(1+\frac{1-s}{s}(\psi(\alpha))^{\kappa n})} < \frac{1}{2}$ at $\beta = 0$. Denote this solution by $b(\alpha)$. The functions $a(\beta)$ and $b(\alpha)$ are continuous and decreasing. It can be easily verified by standard arguments that there is $(\alpha^*, \beta^*) \in (0,1) \times (0,1)$ which solves the two equations.

Since $p(\beta^*(n)) > 2z-1 > 0$ there is $\epsilon > 0$ such that $\beta^*(n) > \epsilon$ for all n. As $\psi(\epsilon) \in (0,1)$, by the second equation $\beta^*(n)$ converges to $\bar{\beta}$, the unique solution of $p(\beta) = 2z-1$ and $\beta^*(n) > \bar{\beta}$. It follows that the probability that all Avg referees will vote Y is less than $s(1-F(\bar{\beta}))^{n(1-\kappa)} + (1-s)(1-G(\bar{\beta}))^{n(1-\kappa)}$ which converges to zero as $n \to \infty$. □

8. Non-Bayesian games: aggregating n signals

In this section, we discuss an alternative modelling approach to belief formation with non-Bayesian procedures. Instead of aggregating two signals, each referee aggregates n signals: his own signal and one distinct signal for each Y vote cast by the other referees. More precisely, a referee i who receives the signal t and knows that all the other referees follow a common cutoff θ forms his C-belief by aggregating:

(i) his own signal t (according to which the defendant is guilty with probability $p(t)$); and

(ii) $n-1$ signals, one for each referee $j \neq i$ who is voting Y, that is, referee j's signal is at least θ (for any one of these signals the defendant is guilty with probability $r^1(\theta)$).

Thus, under the Avg approach, $\mu_i(t,\theta) = \frac{1}{n}(p(t) + (n-1)r^1(\theta))$ and under the Max approach, $\mu_i(t,\theta) = max\{p(t), r^1(\theta)\}$.

This approach makes sense if a referee receives information about the proportion of defendants who are guilty when referees vote Y. This corresponds to a scenario in which information such as "referee j was right in 70% of the cases in which he voted to convict" is available.

We will now see that the conclusions obtained in the previous section essentially remain valid under this approach to belief formation.

Claim D *Suppose each referee aggregates n signals.*

(a) If all referees follow the Avg procedure, then:

(i) There is a unique equilibrium $\bar{\beta}$ satisfying $\frac{1}{n}(p(\bar{\beta}) + (n-1)r^1(\bar{\beta})) = z$.

(ii) The sequence of equilibria $\bar{\beta}(n)$ decreases and converges to $\hat{\beta}$, the unique solution of the equation $r^1(\beta) = z$. If $z > s$, then $\hat{\beta} > 0$ and the probability of conviction converges to 0.

(b) If all referees follow the Max procedure, then:

(i) There exists a unique equilibrium $\bar{\gamma}$ satisfying $r^1(\bar{\gamma}) = z$.

(ii) For all n, $\bar{\gamma} < \bar{\beta}(n)$.

(c) Suppose that $s < \frac{1}{2} < z$, that $n\kappa$ players are Bayesian and that the remaining players are Avg, where $\kappa < 1$. Then:

(i) If $1 - \kappa + \kappa s > z$, then an equilibrium exists when n is sufficiently large. Furthermore, any sequence of equilibria $(\bar{\alpha}(n), \bar{\beta}(n))$ converges to $(0, \beta^*)$ where $(1-\kappa) r^1(\beta^*) + \kappa s = z$ and the probability of conviction goes to zero.

(ii) If $1 - \kappa + \kappa s < z$, then no equilibrium exists for n sufficiently large.[5]

Proof. (a) (i) The function $\frac{1}{n}(p(\beta) + (n-1) r^1(\beta))$ is increasing continuously in β and ranges from $\frac{n-1}{n} s$ to 1. Since $s \leq z < 1$, there is a unique $\bar{\beta} \in (0,1)$ such that the above function equals z. This is the unique equilibrium.
(ii) Since $p(\beta) < r^1(\beta)$ the equilibrium sequence $\bar{\beta}(n)$ decreases and converges to $\hat{\beta}$. When $z > s$, $\hat{\beta} > 0$ and thus the probability of conviction converges to zero as $n \to \infty$.

(b) (i) Since $r^1(t) \geq p(t)$, an equilibrium is characterized by the equation $r^1(\gamma) = z$ for all n. Since $z \geq s$, a solution $\bar{\gamma}$ exists and is unique.
(ii) Recall that $\frac{1}{n}[p(\bar{\beta}(n)) + (n-1) r^1(\bar{\beta}(n))] = z$. Since $p(t) < r^1(t)$ for all $t \in (0,1)$ and $\bar{\beta}(n) \in (0,1)$, it must be that $r^1(\bar{\beta}(n)) > z$ and therefore $\bar{\beta}(n) > \bar{\gamma}$.

(c)(i) The conditions for an equilibrium $(\alpha, \beta) \in (0,1) \times (0,1)$ are:

$$\frac{1}{1 + \frac{g(\alpha)}{f(\alpha)} \frac{1-s}{s} (\psi(\alpha))^{\kappa n - 1} (\psi(\beta))^{n - \kappa n}} = z \text{ and}$$

$$\frac{1}{n}[p(\beta) + (n - n\kappa - 1) r^1(\beta) + n\kappa r^1(\alpha)] = z$$

where, as before, $\psi(t) = \frac{1-G(t)}{1-F(t)}$. Given any $\beta < 1$, let $a(\beta, n)$ be the solution of the first equation. To see that such a solution always exists, note that the LHS converges to 1 as $\alpha \to 1$ and to 0 as $\alpha \to 0$. Also note that $a(\beta, n)$ is decreasing in β, $a(\beta, n) \to 0$ as $\beta \to 1$, and, for any $\beta < 1$, $a(\beta, n) \to 0$ as $n \to \infty$. When n is large enough there is at most one value $b(n)$ that solves the second equation for $\alpha = a(\beta, n)$. The LHS at $\beta = 0$

[5] Recall that interior cutoffs are part of our definition of equilibrium. Naturally, degenerate equilibria always exist.

is $\frac{(n-n\kappa-1)s}{n}+\kappa r^1(a(0,n))$ which converges to $s<z$ as $n\to\infty$. The LHS converges to $1-\kappa+\kappa s>z$ as $\beta\to 1$. Therefore, $(\bar{\alpha}(n),\bar{\beta}(n))$, where $\bar{\alpha}(n)=a(b(n),n)$ and $\bar{\beta}(n)=b(n)$, is an equilibrium.

If there is a subsequence of $\bar{\alpha}(n)$ exceeding some $\epsilon>0$ then, since $\psi(\epsilon)<1$ and $\psi(\beta)\le 1$, the LHS of the first equation converges to $1>z$. Thus, $\bar{\alpha}(n)\to 0$ and $\bar{\beta}(n)$ converges to the unique solution of the equation $(1-\kappa)r^1(\beta)+\kappa s=z$. This solution is positive and therefore the probability of conviction goes to 0.
(ii) Consider a sequence of equilibria $(\bar{\alpha}(n),\bar{\beta}(n))$. Since $a(0,n)\to 0$ as $n\to\infty$ and $a(\bar{\beta}(n),n)\le a(0,n)$ then we have $a(\bar{\beta}(n),n)\to 0$. Thus, the limit superior of the LHS of the second equilibrium condition evaluated at $(\bar{\alpha}(n),\bar{\beta}(n))$ is bounded from above by $(1-\kappa)+\kappa s<z$, contradicting the existence of such a sequence. □

Example: Since a fully Bayesian panel is welfare-maximizing, it can be concluded that having more Bayesian agents in a mixed panel increases the level of welfare. But this is not always true. Suppose that $f(t)=2t$, $g(t)=2-2t$ (with cdfs $F(t)=t^2$ and $G(t)=2t-t^2$ respectively), $s=\frac{1}{2}$ and $z=\frac{1}{2}$. Then, $p(t)=t$ and $r^k(t)=\frac{(1+t)^k}{(1+t)^k+(1-t)^k}$. The equilibrium of the Avg game is the solution to the equation $\frac{1}{n}(\bar{\beta}+(n-1)\frac{1+\bar{\beta}}{2})=\frac{1}{2}$, that is $\bar{\beta}=\frac{1}{n+1}$ (which can be shown to be lower than σ^*, which is the solution to the equation $(\frac{1+\alpha}{1-\alpha})^{n-1}=\frac{1-\alpha}{\alpha}$ whenever $n\ge 3$; for $n=2$, the two solutions $\bar{\beta}$ and σ^* are equal to $\frac{1}{3}$). The equilibrium welfare sequence is $\frac{1}{2}(1-\frac{1}{(n+1)^2})^n+\frac{1}{2}(1-(1-\frac{1}{n+1})^{2n})$ (which converges to $1-\frac{1}{2e^2}=0.932$ as $n\to\infty$).

A panel consisting of only Avg referees can be preferable to a panel of equal size with half of its referees being Bayesian and half being Avg.[6] For instance, the welfare generated by a panel with 14 Avg referees is 0.897 while that generated by an equally sized mixed panel is 0.875.

[6]The equilibrium level of welfare in the case of such mixed panel is $\frac{1}{2}[(1-\alpha^2)^{n/2}(1-\beta^2)^{n/2}+1-(1-2\alpha+\alpha^2)^{n/2}(1-2\beta+\beta^2)^{n/2}]$.

References

Bar-Hillel, Maya (1980). "The base-rate fallacy in probability judgments". *Acta Psychologica, 44*(3), 211-233.

Duggan, John and César Martinelli (2001). "A Bayesian model of voting in juries". *Games and Economic Behavior, 37*(2), 259-294.

Kahneman, Daniel and Amos Tversky (1973). "On the psychology of prediction". *Psychological Review, 80*(4), 237-251.

Porat, Ariel and Eric Posner (2012). "Aggregation and law". *Yale Law Journal*, 122(1), 2-69.