Can Words Get in the Way? Supplemental Appendix

Matias Iaryczower
Princeton

Xiaoxia Shi
UW-Madison

Matthew Shum
Caltech*

August 31, 2016

*Department of Politics, Princeton University, Princeton NJ 08540; Department of Economics, University of Wisconsin- Madison, 1180 Observatory Drive, Madison, WI 53706; and Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125. Emails: miaryc@princeton.edu, xshi@ssc.wisc.edu, and mshum@caltech.edu.

A Appendix: Additional Figures and Tables

Table A.1: Summary statistics of data variables

Variable:		Mean	Std.Dev.
Case characteristics:			
FedLaw	=1 if case prosecuted under federal law	0.8169	0.3868
Aggravated	=1 if crime is aggravated assault/murder	0.1220	0.3273
White Collar	=1 if white collar crime	0.2038	0.4029
Theft	=1 if crime is theft	0.1414	0.3485
Narcotics	=1 if drug-related crime	0.2062	0.4047
Rep. Majority	=1 if ≥ 2 republicans on panel	0.4452	0.4971
Female	=1 if ≥ 1 female judge on panel	0.0831	0.2760
Harv-Yale Majority	=1 if ≥ 2 Harvard/Yale grads on panel	0.1812	0.3853
Jury instruction	=1 if main legal issue is jury instruction	0.1970	0.3978
Sentencing	=1 if main legal issue is sentencing	0.1624	0.3689
Admissibility	=1 if main legal issue is admissibility of evidence	0.3473	0.4762
Sufficiency	=1 if main legal issue is sufficiency of evidence	0.2547	0.4358
# cases:		3239	
Judge characteristics:			
Republican	=1 if judge is republican	0.5392	0.4989
Yearsexp	Years of experience	7.3174	7.1620
Judexp	Years of prior judicial experience	7.1893	7.8409
Polexp	Years of prior political experience	1.9197	3.7628
#judges:		523	
Vote Outcomes:			
Unanimous to Overturn		21.0%	
Divided to Overturn		2.8%	
Divided to Uphold		4.0%	
Unanimous to Uphold		72.2%	

Table A.2: Average Case and Judge Characteristics Across Vote Outcomes

	Unanimous 1	Unanimous 0	Divided 1	Divided 0
Case Specific:				
FedLaw	0.787	0.832	0.750	0.754
Aggravated	0.124	0.114	0.185	0.208
White Collar	0.179	0.209	0.207	0.238
Theft	0.149	0.137	0.163	0.169
Narcotics	0.166	0.222	0.174	0.154
Rep. Majority	0.446	0.451	0.391	0.377
Female in Panel	0.081	0.082	0.120	0.092
Harvard-Yale Majority	0.168	0.186	0.185	0.169
Jury Instruction	0.168	0.206	0.098	0.254
Sentencing	0.138	0.173	0.087	0.146
Admissibility	0.281	0.362	0.326	0.446
Sufficiency	0.178	0.281	0.196	0.231
Caseload	42.835	44.160	38.357	39.945
Judge Specific:				
Republican	0.456	0.472	0.399	0.444
Year of Experience	9.535	9.664	10.210	10.118
Prior Judicial Experience	5.518	5.493	4.902	4.862
Prior Political Experience	1.904	1.825	1.645	2.013
Rep×Assault	0.056	0.053	0.087	0.100
Rep×White Collar	0.085	0.095	0.080	0.115
Rep×Theft	0.058	0.060	0.062	0.044
Rep×Narcotics	0.082	0.118	0.065	0.074
Nonwhite Dummy	0.048	0.040	0.029	0.064
Female Dummy	0.027	0.029	0.040	0.033
Number of Cases:	680	2337	92	130

Note: Unanimous 1: Unanimous Overturn; Unanimous 0: Unanimous Uphold; Divided 1; Divided Overturn; Divided 0; Divided Uphold.

Table A.3: Benchmark specification

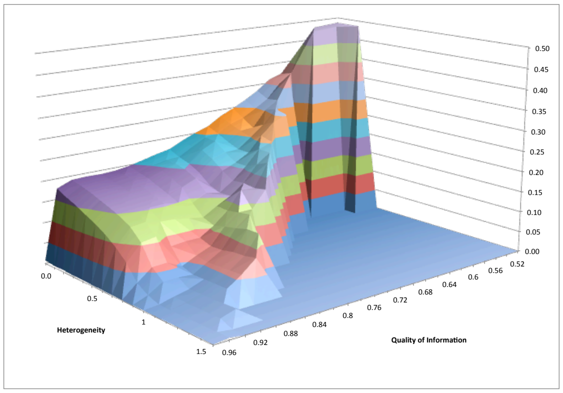
Estimated Vote Probabilities $p_v(\vec{v}\|X)$:			
$\hat{p}_v(111)$	=0.224	$\hat{p}_v(000)$	=0.676
$\hat{p}_v(101)$	=0.020	$\hat{p}_v(010)$	=0.015
$\hat{p}_v(110)$	=0.012	$\hat{p}_v(001)$	=0.019
$\hat{p}_v(100)$	=0.025	$\hat{p}_v(011)$	=0.010

Case characteristics:			
FedLaw	=1	Jury instruction	=0
Narcotics	=0	Sentencing	=0
Aggravated	=0	Admissibility	=1
White Collar	=1	Sufficiency	=0
Theft	=0	Rep. Majority	=1
Female Judge	=0	Harvard-Yale Majority	=0

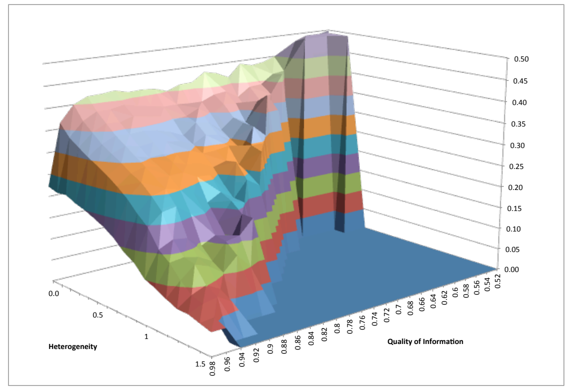
Judge characteristics:	Judge 1	Judge 2	Judge 3
Republican	1	1	0
Yearsexp	7.19	0	7.19
Judexp	1.92	0	1.92
Polexp	0	6.85	6.85

Table A.4: Average Judge Characteristics for Dissenting Judges and Non-dissenting Judges

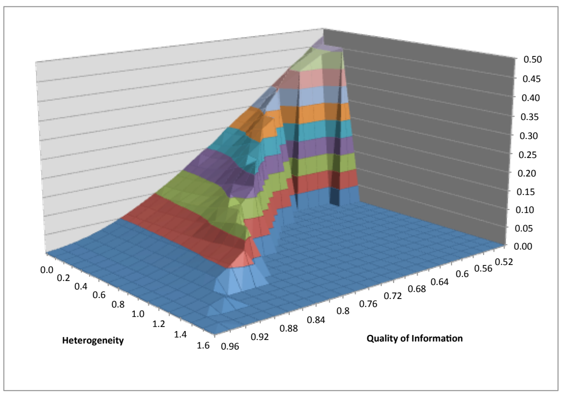
	Dissenting	Non-dissenting	Difference	Standard Error
Republican	0.419	0.466	-0.047	0.034
Years of Experience	10.554	9.650	0.904	0.515
Prior Judicial Experience	5.027	5.465	-0.439	0.496
Prior Political Experience	1.500	1.852	-0.352	0.234
Rep*Assault	0.086	0.056	0.030	0.019
Rep*White Collar	0.104	0.093	0.010	0.021
Rep*Theft	0.059	0.059	-0.000	0.016
Rep*Narcotics	0.063	0.108	-0.045	0.017
Nonwhite Dummy	0.072	0.042	0.030	0.018
Female Dummy	0.041	0.029	0.011	0.013



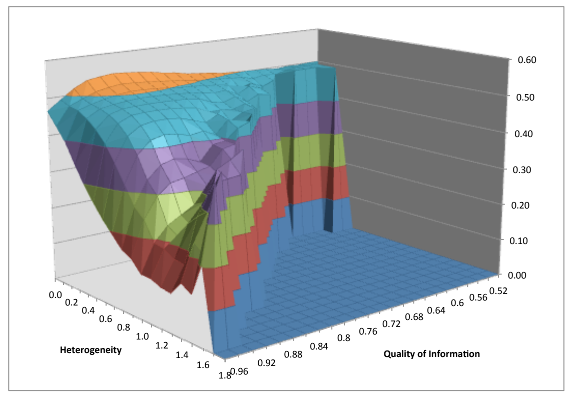
Min. Error $\underline{\varepsilon}^*(\theta, \vec{\pi}, p_v(\vec{v}))$ - Eq. c.w./ data



Max. Error $\overline{\varepsilon}^*(\theta, \vec{\pi}, p_v(\vec{v}))$ - Eq. c.w./ data

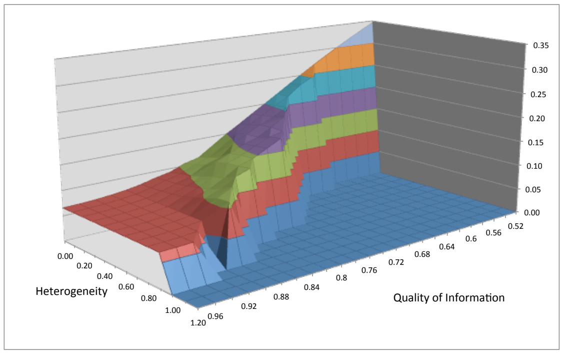


Min. Error $\underline{\varepsilon}(\rho, q, \pi)$ - All Equilibria

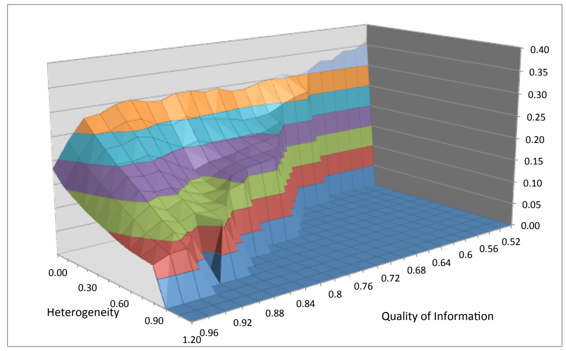


Max. Error $\overline{\varepsilon}(\rho, q, \pi)$ - All Equilibria

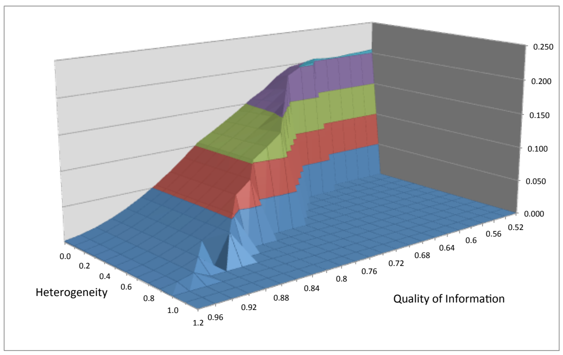
Figure A.1: Minimum and maximum probability of error in equilibria consistent with the data (top) and all equilibria (bottom), for pairs of preference heterogeneity and competence (H, q) consistent with points in the confidence set for $\rho = 0.5$. (Average of extrema across points $(\vec{\pi}, q)$ such that $H(\vec{\pi}) = H$).



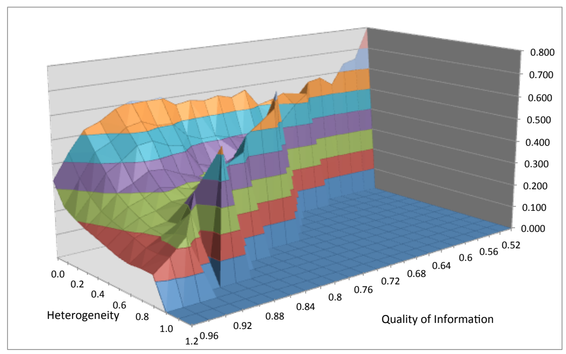
Min. Error $\underline{\varepsilon}^*(\theta, \vec{\pi}, p_v(\vec{v}))$ - Eq. c.w./data



Max. Error $\overline{\varepsilon}^*(\theta, \vec{\pi}, p_v(\vec{v}))$ - Eq. c.w./data

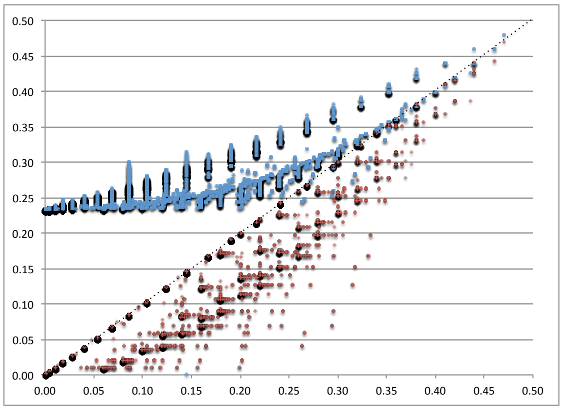


Min. Error $\underline{\varepsilon}(\rho, q, \pi)$ - All Equilibria

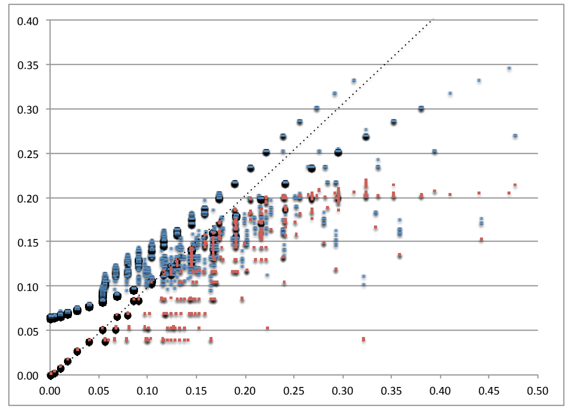


Max. Error $\overline{\varepsilon}(\rho, q, \pi)$ - All Equilibria

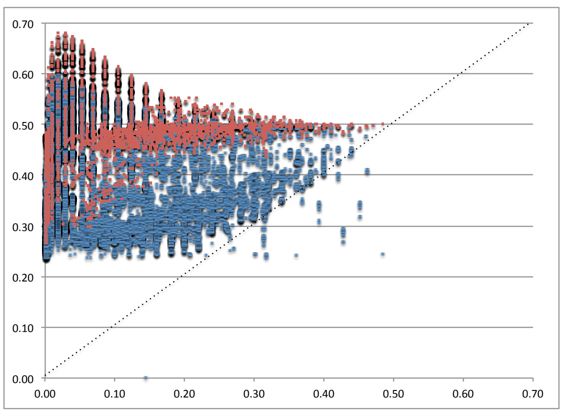
Figure A.2: Minimum and maximum equilibrium probability of error in equilibria consistent with the data (top) and all equilibria (bottom), for pairs of preference heterogeneity and competence (H, q) consistent with points in the EIS for $\rho = 0.2$. (Average of extrema across points $(\vec{\pi}, q)$ such that $H(\vec{\pi}) = H$).



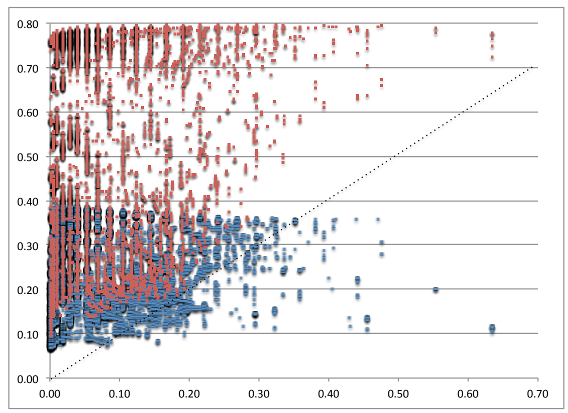
Min. Error ($\rho = 0.5$)



Min. Error ($\rho = 0.2$)

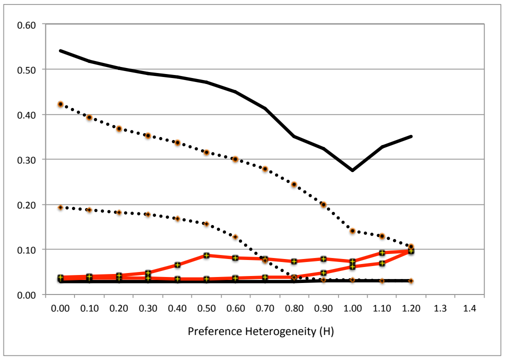


Max. Error ($\rho = 0.5$)

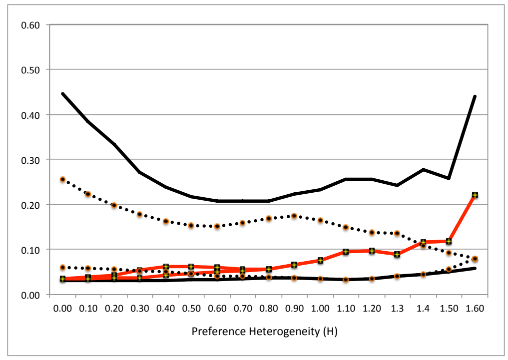


Max. Error ($\rho = 0.2$)

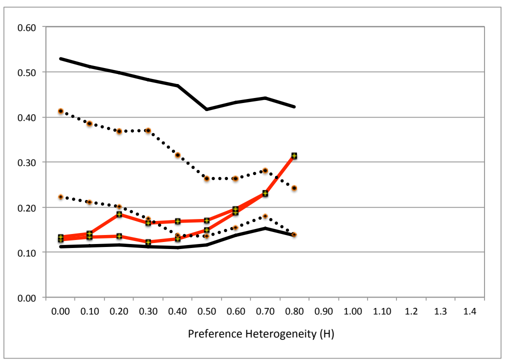
Figure A.3: For each point in the EIS, red dots plot the correspondence between the min/max probability of error in equilibria without deliberation (x-axis) and the min/max probability of error in all equilibria with deliberation (y-axis). Blue dots plot the correspondence between min/max probability of error in equilibria without deliberation (x-axis) and the min/max probability of error in equilibria with deliberation consistent with the data.



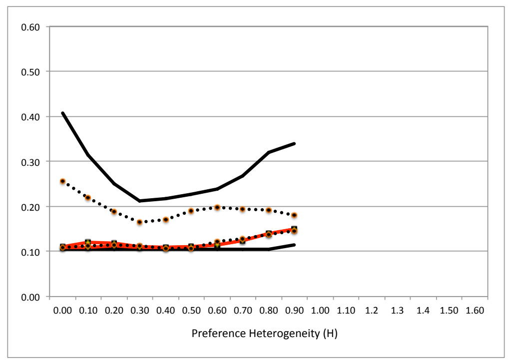
$(\rho = 0.5; q = 0.90)$



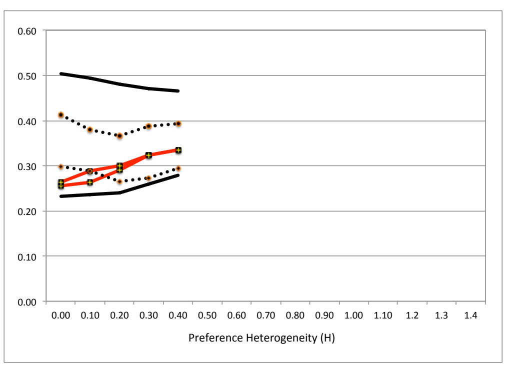
$(\rho = 0.2; q = 0.90)$



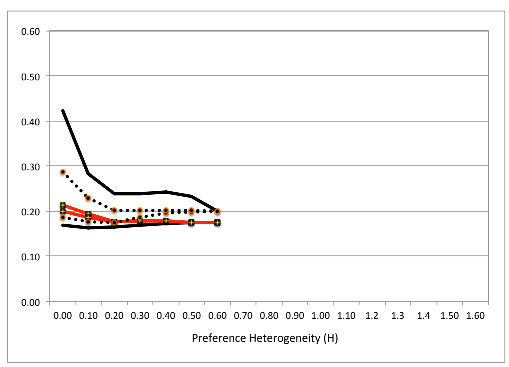
$(\rho = 0.5; q = 0.80)$



$(\rho = 0.2; q = 0.80)$



$(\rho = 0.5; q = 0.70)$



$(\rho = 0.2; q = 0.70)$

Figure A.4: Probability of mistakes with and without deliberation for values of preference heterogeneity H consistent with points $(\vec{\pi}, \theta)$ in the confidence set for $\rho = 0.5$ (left), and $\rho = 0.2$ (right). Min. and max. eq. probability of error in (i) all equilibria with deliberation (solid black), (ii) in equilibria with deliberation consistent with the data (dotted), and (iii) in responsive equilibria without deliberation (solid red, with marker).

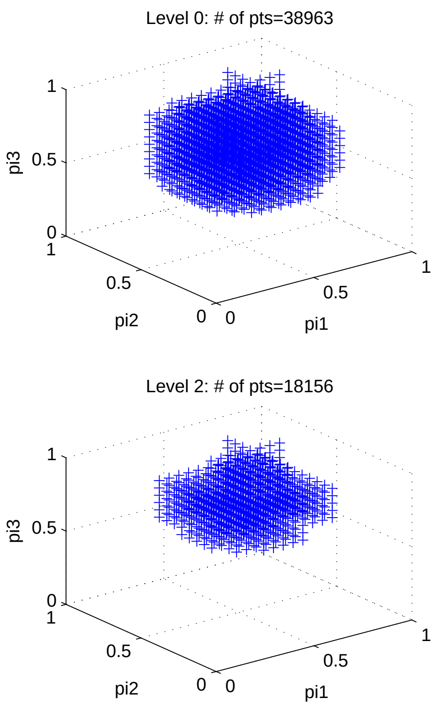
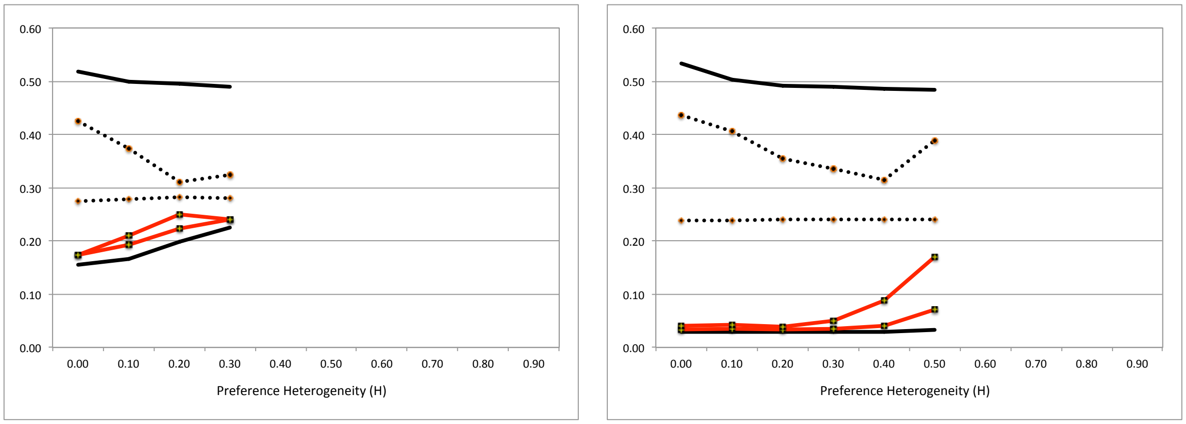
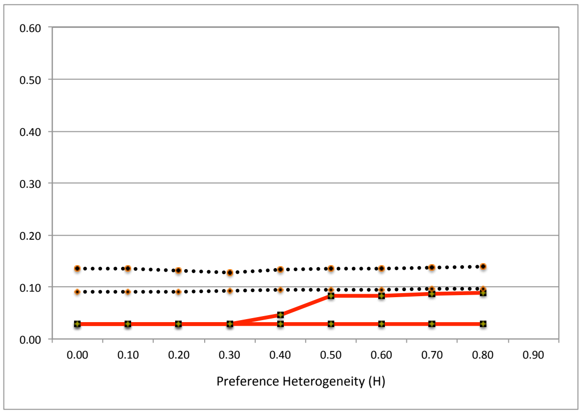


Figure A.5: Estimated Identified set for (π_1, π_2, π_3) with $\rho = 0.5$ and $q = 0.76$. Level 0: the estimated identified set using the original constraints in the paper under the benchmark specification (38,963 points). Level 1: intersected with 0.5 quantile of CASELOAD; Level 2: intersected with 0.25, 0.5, 0.75 quantiles of CASELOAD; Level 3: intersected with 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875 quantiles of CASELOAD.

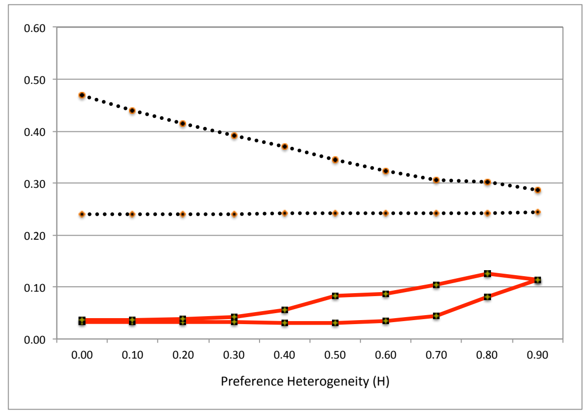


($\rho = 0.5; q = 0.76$) ($\rho = 0.5; q = 0.90$)

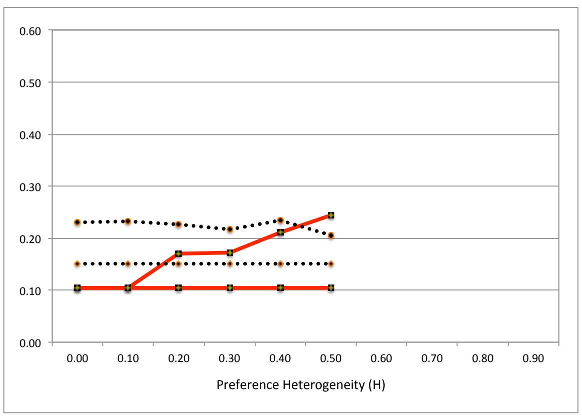
Figure A.6: Probability of mistakes with and without deliberation for values of preference heterogeneity H consistent with points $(\vec{\pi}, \theta)$ in the refined EIS for $\rho = 0.5$. Min. and max. eq. probability of error in (i) all equilibria with deliberation (solid black), (ii) in equilibria with deliberation consistent with the data (dotted), and (iii) in responsive equilibria without deliberation (solid red, with marker).



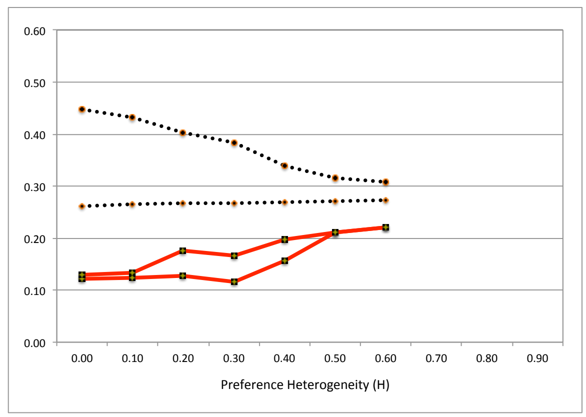
$(\hat{p}^{\text{dissent}}; \rho = 0.5; q = 0.90)$



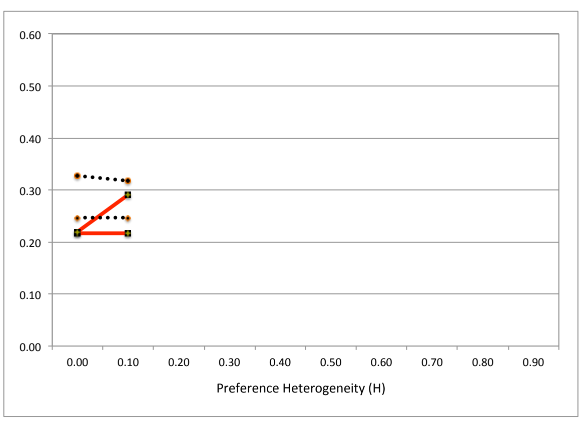
$(\hat{p}; \rho = 0.5; q = 0.90)$



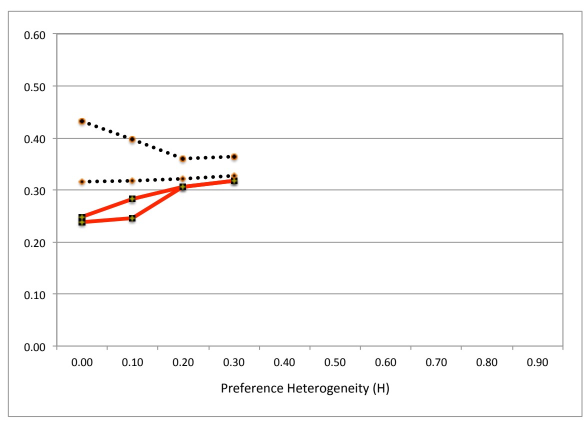
$(\hat{p}^{\text{dissent}}; \rho = 0.5; q = 0.80)$



$(\hat{p}; \rho = 0.5; q = 0.80)$



$(\hat{p}^{\text{dissent}}; \rho = 0.5; q = 0.70)$



$(\hat{p}; \rho = 0.5; q = 0.70)$

Figure A.7: Probability of mistakes with and without deliberation for values of preference heterogeneity H consistent with points $(\vec{\pi}, \theta)$ in the EIS for $\rho = 0.5$ (left), based on (left) an artificial dataset in which all unanimous cases have been discarded and (right) based on actual data

B Confidence Set

Here, we discuss statistical inference in partially identified models based on confidence sets which cover either the true parameter, or the identified set with a pre-specified probability. Following the literature, we construct a confidence set by inverting a test for the null hypothesis $H_0 : (\theta, \vec{\pi}) \in \mathcal{A}_0$ for each fixed $(\theta, \vec{\pi})$. To be specific, we collect all the $(\theta, \vec{\pi})$ such that there is one $\mu \in M(\theta, \vec{\pi})$ at which the H_0 is not rejected. The collection of all those $(\theta, \vec{\pi})$ forms a confidence set.[1]

Standard application of the central limit theorem gives us $\sqrt{n}(\hat{\vec{p}}_v - \vec{p}_v) \to_d N(0, \Sigma)$, where Σ is the variance-covariance matrix of the vector of vote probabilities $\vec{p}_v$. Then the law of large numbers implies $\hat{\Sigma}_n \to_p \Sigma$. Accordingly, we define the following test statistic:

$$T_n(\theta, \vec{\pi}) = nQ_n(\theta, \vec{\pi}; \hat{\Sigma}_n^{-1}). \tag{B.1}$$

By definition, $T_n(\theta, \vec{\pi}) \leq nQ_n(\theta, \vec{\pi}, \mu; \hat{\Sigma}_n^{-1})$ for any $(\theta, \vec{\pi}, \mu) \in \mathcal{B}_0$. Using standard arguments, we can show that for any $(\theta, \vec{\pi}, \mu) \in \mathcal{B}_0$, $nQ_n(\theta, \vec{\pi}, \mu, \hat{\Sigma}_n^{-1}) \to_d \chi^2(7)$. Thus, a test of significance level $\alpha \in (0, 1)$ can use the $1 - \alpha$ quantile of $\chi^2(7)$ as critical value. The confidence set for $(\theta, \vec{\pi})$ is defined as

$$CS_n(1 - \alpha) = \{(\theta, \vec{\pi}) \in \Theta \times [0, 1]^3 : T_n(\theta, \vec{\pi}) \leq \chi^2_{7,\alpha}\}, \tag{B.2}$$

where $\chi^2_{7,\alpha}$ is the $1 - \alpha$ quantile of $\chi^2(7)$.

Theorem 2. *Suppose Σ is invertible. Then*
(a) $\liminf_{n\to\infty} \inf_{(\theta,\vec{\pi})\in\mathcal{A}_0} \Pr((\theta, \vec{\pi}) \in CS_n(1 - \alpha)) \geq 1 - \alpha$; *and*
(b) $\liminf_{n\to\infty} \Pr(\mathcal{A}_0 \subseteq CS_n(1 - \alpha)) \geq 1 - \alpha$.

Proof. See Appendix C. □

Remark B.1. Part (a) shows that CS_n covers the true value of $(\theta, \vec{\pi})$ with asymptotic probability no smaller than $1 - \alpha$. Interestingly, it is also a confidence set that covers $\mathcal{A}_0$ with asymptotic probability no smaller than $1 - \alpha$, as shown in part (b).[2] The intuition for this phenomenon is that the random components of $T_n(\theta, \vec{\pi}, \mu)$ – which are just the empirical frequencies of the vote probabilities $\hat{\vec{p}}$ – do not depend on the model parameters (θ, π).

[1] This inferential method differs from the approach of Pakes, Porter, Ho, and Ishii (2015), which is based on moment inequalities derived from agents' best-response correspondences. While this approach has proved useful in several applications with games of complete information, in the context of our incomplete information environment we have not been able to derive moment inequalities based on best-response behavior.

[2] Imbens and Manski (2004) initiated a sizable literature regarding these two types of confidence sets.

In contrast, in typical moment inequality models, the random sample moment functions depend explicitly on the model parameters.

Remark B.2. Because the confidence set CS_n above is based on the asymptotic critical value for $nQ_n(\theta, \vec{\pi}, \mu; \hat{\Sigma}_n^{-1})$, which is weakly bigger than $T_n(\theta, \vec{\pi})$, it may over-cover asymptotically; that is, it may be larger than necessary. Tighter and nonconservative confidence sets can be constructed by directly approximating the distribution of $T_n(\theta, \vec{\pi})$ using the methods developed in Bugni, Canay, and Shi (2013) and Kitamura and Stoye (2011).[3] The disadvantage of doing this is two-fold: (i) the critical value will need to be simulated and will depend on θ and $\vec{\pi}$; and (ii) a tuning parameter will need to be introduced to reflect the slackness of the inequality constraints. In addition, in our data, we find that the confidence set CS_n is not much larger than the EIS $\hat{\mathcal{A}}_n$, suggesting that not much can be gained by adopting the more complicated methods.

The confidence set can be computed in the following steps:

(1) for each $(\theta, \vec{\pi})$, compute $T_n(\theta, \vec{\pi}) = nQ_n(\theta, \vec{\pi}; \hat{\Sigma}_n^{-1})$ via the quadratic program:

$$\begin{aligned} Q_n(\theta, \vec{\pi}; W_n) = & \min_{\vec{\mu} \in [0,1]^{64}} (\vec{p}_v - P_t(\theta)\vec{\mu})' W (\vec{p}_v - P_t(\theta)\vec{\mu})' \\ & s.t. (3.2), (3.3), (3.4), \text{ and } \sum_{j=k+1}^{k+8} \vec{\mu}_j = 1, \ k = 0, ..., 7. \end{aligned} \tag{B.3}$$

(2) repeat step (1) for many grid points of $(\theta, \vec{\pi}) \in \Theta \times [0,1]^3$, and

(3) collect the points in step (2) that satisfy $T_n(\theta, \vec{\pi}) \leq \chi^2_{7,\alpha}$, and the points form $CS_n(1-\alpha)$.

For all the results in this paper, we use a value of $\alpha = 0.05$.

[3] See Wolak (1989) for the case where the inequality constraints are linear in the structural parameters θ.

C Proofs

Proof of Theorem 1. Because $p(\vec{t},\theta) = p(\vec{t}|w=1;\theta)\rho + p(\vec{t}|w=0;\theta)(1-\rho)$ is continuously differentiable in θ, Theorem 2.1 of Shi and Shum (2015) applies and shows that $d_H(\hat{\mathcal{B}}_n, \mathcal{B}_0) \to_p 0$, where

$$\hat{\mathcal{B}}_n = \{(\theta,\vec{\pi},\vec{\mu}) \in \mathcal{B} : Q_n(\theta,\vec{\pi},\mu;W_n) = \min_{(\theta,\vec{\pi})\in\Theta\times[0,1]^3} Q_n(\theta,\vec{\pi},W_n)\}, \tag{C.1}$$

where $Q_n(\theta,\vec{\pi},\mu;W_n)$ is defined like $Q(\theta,\vec{\pi},\mu;W)$ but with $\vec{p}$ and W replaced by $\hat{\vec{p}}$ and W_n. Because $\hat{\mathcal{A}}_n$ and $\mathcal{A}_0$ are the projections of $\hat{\mathcal{B}}_n$ and $\mathcal{B}_0$ onto their first $d_\theta + 3$ dimension, respectively, we have $d_H(\hat{\mathcal{A}}_n, \mathcal{A}_0) \to_p 0$. □

Proof of Theorem 2. (a) For any sequence $\{(\theta_n,\vec{\pi}_n) \in \mathcal{A}_0\}_{n=1}^\infty$, there exists $\{\mu_n \in M(\theta_n,\vec{\pi}_n)\}_{n=1}^\infty$ such that $\vec{p}_v = P_t(\theta_n)\vec{\mu}_n$. Thus, $nQ_n(\theta_n,\vec{\pi}_n,\mu_n;\hat{\Sigma}_n^{-1}) = n(\hat{\vec{p}}_v - \vec{p}_v)'\hat{\Sigma}_n^{-1}(\hat{\vec{p}}_v - \vec{p}_v) \to_d \mathcal{X}^2(7)$. Thus

$$\begin{aligned}
\Pr((\theta_n,\vec{\pi}_n) \in CS_n(1-\alpha)) =& \Pr(T_n(\theta_n,\vec{\pi}_n) \le \chi^2_{7,\alpha}) \\
&\ge \Pr(nQ_n(\theta_n,\vec{\pi}_n,\mu_n;\hat{\Sigma}_n^{-1}) \le \chi^2_{7,\alpha}) \\
&\to \Pr(\chi^2(7) \le \chi^2_{7,\alpha}) = 1-\alpha.
\end{aligned} \tag{C.2}$$

This implies part (a).

(b) Part (b) holds because

$$\begin{aligned}
\Pr(\mathcal{A}_0 \subseteq CS_n(1-\alpha)) &= \Pr(\sup_{(\theta,\vec{\pi})\in\mathcal{A}_0} T_n(\theta,\vec{\pi}) \le \chi^2_{7,\alpha}) \\
&\ge \Pr(\sup_{(\theta,\vec{\pi},\mu)\in\mathcal{B}_0} nQ_n(\theta,\vec{\pi},\mu;\hat{\Sigma}_n^{-1}) \le \chi^2_{7,\alpha}) \\
&= \Pr(n(\hat{\vec{p}}_v - \vec{p}_v)'\hat{\Sigma}_n^{-1}(\hat{\vec{p}}_v - \vec{p}_v) \le \chi^2_{7,\alpha}) \\
&\to \Pr(\chi^2(7) \le \chi^2_{7,\alpha}) = 1-\alpha,
\end{aligned} \tag{C.3}$$

where the second equality holds because for all $(\theta,\vec{\pi},\mu) \in \mathcal{B}_0$, $\vec{p}_v = P_t(\theta)\vec{\mu}$. □

D Responsive Equilibria without Deliberation

In Section 5.4 in the main paper we compared the equilibrium probability of error in voting with deliberation with the corresponding equilibrium probability of error that would have occurred in the absence of deliberation for the same court and case characteristics. Specifically, for each point $(\theta, \vec{\pi})$ in the confidence set we compare the maximum and minimum error probabilities across all equilibria, $\overline{\varepsilon}(\theta, \vec{\pi})$ and $\underline{\varepsilon}(\theta, \vec{\pi})$, and across equilibria consistent with the data, $\overline{\varepsilon}^*(\theta, \vec{\pi}, p_v)$ and $\underline{\varepsilon}^*(\theta, \vec{\pi}, p_v)$, with the corresponding maximum and minimum error probabilities in responsive Bayesian Nash equilibria (BNE) of the voting game *without* communication, $\overline{\varepsilon}^{ND}(\theta, \vec{\pi})$ and $\underline{\varepsilon}^{ND}(\theta, \vec{\pi})$. To carry out this comparison, we solve for all responsive BNE of the non-deliberation game, for all parameter points in the confidence set.

In the game without deliberation, the strategy of player i is a mapping $\sigma_i : \{0, 1\} \rightarrow [0, 1]$, where $\sigma_i(t_i)$ denotes the probability of voting to overturn given signal t_i. It is easy to show that $\sigma_i(t_i) > 0$ (< 1) only if $\Pr(\omega = 1 | t_i, Piv^i) \geq \pi_i$ $(\leq \pi_i)$, or

$$\frac{\Pr(t_i|\omega = 1)}{\Pr(t_i|\omega = 0)} \frac{\Pr(Piv^i|\omega = 1)}{\Pr(Piv^i|\omega = 0)} \geq \frac{\pi_i}{1 - \pi_i} \frac{1 - \rho}{\rho} \tag{D.1}$$

Let $\alpha_{i\omega} \equiv \Pr(v_i = 1|\omega)$ denote the conditional probability that i votes to overturn in state ω, and note that $\alpha_{i1} = q_i\sigma_i(1) + (1 - q_i)\sigma_i(0)$, and $\alpha_{i0} = (1 - q_i)\sigma_i(1) + q_i\sigma_i(0)$. Substituting in (D.1), we have that $\sigma_i(t_i) > 0$ only if (for $j, k \neq i$)

$$\frac{\Pr(t_i|\omega = 1)}{\Pr(t_i|\omega = 0)} \left[\frac{\alpha_{j1}(1 - \alpha_{k1}) + \alpha_{k1}(1 - \alpha_{j1})}{\alpha_{j0}(1 - \alpha_{k0}) + \alpha_{k0}(1 - \alpha_{j0})}\right] \geq \frac{\pi_i}{1 - \pi_i} \frac{1 - \rho}{\rho} \tag{D.2}$$

Under certain conditions (when the court is sufficiently homogeneous) there is an equilibrium in which all judges vote *informatively*; i.e., $\sigma_i(1) = 1, \sigma_i(0) = 0$ for all $i \in N$. Note that with informative voting $\alpha_{i1} = q_i$, and $\alpha_{i0} = (1 - q_i)$. Then informative voting is a best response for each i iff

$$\frac{\rho(1 - q_i)}{\rho(1 - q_i) + (1 - \rho)q_i} \leq \pi_i \leq \frac{\rho q_i}{\rho q_i + (1 - \rho)(1 - q_i)}$$

In general, other responsive equilibria are possible. With binary signals and a symmetric environment ($q_i = q$ and $\pi_i = \pi$ $\forall i \in N$), the literature has focused on symmetric responsive BNE. Here the restriction has no bite. Still, there is a relatively "small" class of equilibrium candidates for any given parameter value. The exhaustive list is presented in Table D.5.

Table D.5: Types of Possible Responsive Bayesian Nash Equlibria in the Non-deliberation Game (σ_j in column $\sigma_j(t)$ denotes mixing probability of action t).

	Judge i		Judge j		Judge ℓ		Non-Generic
Eq. Class	$\sigma_1(1)$	$\sigma_1(0)$	$\sigma_2(1)$	$\sigma_2(0)$	$\sigma_3(1)$	$\sigma_3(0)$	
Pure Strategies:							
(EQ1.a)	1	0	1	0	1	0	
(EQ1.b)	1	0	1	0	1	1	
(EQ1.c)	1	0	1	0	0	0	
All judges mix:							
(EQ2)	σ_1	0	σ_2	0	σ_3	0	
(EQ3)	1	σ_1	1	σ_2	1	σ_3	
(EQ4)	σ_1	0	σ_2	0	1	σ_3	
(EQ5)	σ_1	0	1	σ_2	1	σ_3	
Two judges mix:							
(EQ6.a)	σ_1	0	σ_2	0	1	1	
(EQ6.b)	σ_1	0	σ_2	0	0	0	X
(EQ6.c)	σ_1	0	σ_2	0	1	0	
(EQ7.a)	1	σ_1	1	σ_2	1	1	X
(EQ7.b)	1	σ_1	1	σ_2	0	0	
(EQ7.c)	1	σ_1	1	σ_2	1	0	
(EQ8.a)	σ_1	0	1	σ_2	1	1	X
(EQ8.b)	σ_1	0	1	σ_2	0	0	X
(EQ8.c)	σ_1	0	1	σ_2	1	0	

Characterizing responsive equilibria in the non-deliberation game is a laborious but simple task. We illustrate the main logic in case (8.c) in Table D.5; i.e., $\sigma_i(1) \in (0,1), \sigma_j(0) \in (0,1)$, $\sigma_i(0) = 0, \sigma_j(1) = 1$, and $\sigma_k(1) = 1$, $\sigma_k(0) = 0$. (The analysis of the other cases is similar; full details are available upon request). Note that here $\alpha_{10} = (1-q_1)\sigma_1(1)$, $\alpha_{11} = q_1\sigma_1(1)$, $\alpha_{20} = (1-q_2) + q_2\sigma_2(0)$, $\alpha_{21} = q_2 + (1-q_2)\sigma_2(0)$, $\alpha_{30} = 0$, and $\alpha_{31} = 1$.

In equilibrium, $i = 1$ has to be indifferent between upholding and overturning after $t_1 = 1$. Then if it exists, $\sigma_2^*(0)$ is given by the value of $\sigma_2(0) \in [0,1]$ that solves (D.2) with equality for $i = 1$ and $s_i = 1$, or

$$\sigma_2^*(0) = \frac{[q_1(1-\pi_1)\rho - (1-q_1)\pi_1(1-\rho)][(1-q_2)q_3 + q_2(1-q_3)]}{(2q_3-1)[q_1(1-\pi_1)\rho(1-q_2) + (1-q_1)\pi_1(1-\rho)q_2]},$$

which in turn implies $\alpha_{20}^* = (1-q_2) + q_2\sigma_2^*(0)$ and $\alpha_{21}^* = q_2 + (1-q_2)\sigma_2^*(0)$. Similarly, in equilibrium, $i = 2$ has to be indifferent between upholding and overturning after $t_2 = 0$. Then when it exists, $\sigma_1^*(1)$ is given by the value of $\sigma_1(1) \in [0,1]$ that solves (D.2) with equality for $i = 2$ and $t_2 = 0$, or

$$\sigma_1^*(1) = \frac{(1-q_2)q_3(1-\pi_2)\rho - q_2(1-q_3)\pi_2(1-\rho)}{(2q_3-1)[(1-q_2)q_1(1-\pi_2)\rho + q_2(1-q_1)\pi_2(1-\rho)]},$$

which implies $\alpha_{10}^* = (1-q_1)\sigma_1^*(1)$ and $\alpha_{11}^* = q_1\sigma_1^*(1)$. Finally, in equilibrium $i = 3$ has to have incentives to vote informatively. This means that

$$\underbrace{\frac{1-q_3}{q_3}}_{t_3=1} \leq \frac{\alpha_{21}^*(1-\alpha_{11}^*) + \alpha_{11}^*(1-\alpha_{21}^*)}{\alpha_{20}^*(1-\alpha_{10}^*) + \alpha_{10}^*(1-\alpha_{20}^*)} \cdot \frac{1-\pi_3}{\pi_3} \cdot \frac{\rho}{1-\rho} \quad \underbrace{\leq \frac{q_3}{1-q_3}}_{t_3=0}.$$

We can then evaluate numerically, for each point $(\rho, \vec{q}, \vec{\pi})$ in the confidence set, if the conditions for this to be an equilibrium are satisfied. As before, the error associated with this equilibrium σ is $\varepsilon^{ND}(\sigma,\theta) = (1-\rho)\Pr(v = 1|\omega = 0; \sigma, \theta) + \rho \Pr(v = 0|\omega = 1; \sigma, \theta)$, where given majority rule and independent mixing, for $k, \ell \neq j$

$$\Pr(v = 1|\omega, \sigma, \theta) = \sum_{j=1}^{3} \alpha_{k\omega}\alpha_{\ell\omega}(1-\alpha_{j\omega}) + \alpha_{1\omega}\alpha_{2\omega}\alpha_{3\omega}.$$

E Efficient Deliberation

Here we compare social welfare in the equilibria that maximize the sum of judges' payoffs with and without deliberation, for equilibria consistent with the data and all equilibria.

For a given point $(\theta, \vec{\pi})$, and given a communication equilibrium μ, judge i's expected utility is given by the expected cost of type I and type II errors,

$$U_i(\mu; (\theta, \vec{\pi})) = -\left[\rho \varepsilon_{II}(\mu; (\theta, \vec{\pi}))(1 - \pi_i) + (1 - \rho)\varepsilon_I(\mu; (\theta, \vec{\pi}))\pi_i\right].$$

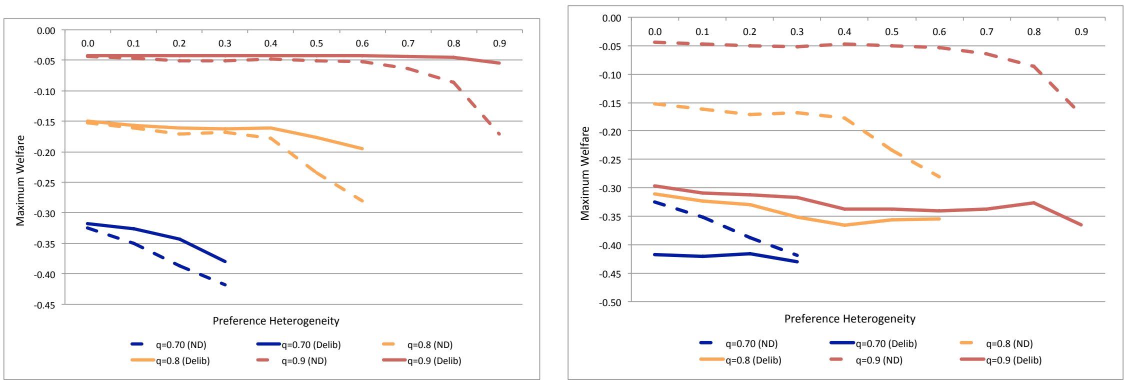
Therefore, the equilibrium that maximizes judges' total welfare, $\mu^*(\theta, \vec{\pi})$, is the $\mu \in M(\theta, \vec{\pi})$ that maximizes $\mathcal{U}(\theta, \vec{\pi}, \mu) \equiv \sum_i U_i(\mu; (\theta, \vec{\pi}))$. A similar definition applies for non-deliberation equilibria, giving $\sigma^*(\theta, \vec{\pi})$. For equilibria consistent with the data, the equilibrium that maximizes judges' total welfare, $\tilde{\mu}(\theta, \vec{\pi})$, is

$$\tilde{\mu}(\theta, \vec{\pi}) = \arg \max_{\mu \in M(\theta, \vec{\pi})} \left\{ \mathcal{U}(\theta, \vec{\pi}, \mu) \quad \text{s.t.} \quad p_v(\vec{v}) = \sum_{\vec{t}} \mu(\vec{v}|\vec{t}) p(\vec{t}; \theta) \right\}.$$

The left panel of Figure E.8 plots the maximum aggregate welfare for points in the EIS for $\rho = 0.5$ across all equilibria of the game with deliberation, $\mathcal{U}^D(\theta, \vec{\pi}) \equiv \mathcal{U}(\mu^*(\theta, \vec{\pi}); (\theta, \vec{\pi}))$, and in the game without deliberation, $\mathcal{U}^N(\theta, \vec{\pi}) \equiv \mathcal{U}(\sigma^*(\theta, \vec{\pi}); (\theta, \vec{\pi}))$. The difference is plotted for various levels of competence q, as a function of the degree of preference heterogeneity in the court. The plot shows that the gain from *efficient deliberation* is fairly small, and concentrated at higher levels of competence and preference heterogeneity. Consider for example the highest competence level plotted in the figure ($q = 0.9$). For all levels of heterogeneity $H \leq 0.8$, the change in aggregate welfare attained by introducing efficient deliberation is smaller than the change in welfare that would result from increasing competence from $q = 0.80$ to $q = 0.90$, or from $q = 0.70$ to $q = 0.80$. Only at $H = 0.9$ is the gain from efficient deliberation relatively high, exceeding the change in welfare that would result from increasing competence from $q = 0.80$ to $q = 0.90$ at low levels of preference heterogeneity.

The right panel provides a similar comparison restricting to the maximum aggregate welfare across equilibria consistent with the data, $\tilde{\mathcal{U}}^D(\theta, \vec{\pi}) \equiv \mathcal{U}(\tilde{\mu}(\theta, \vec{\pi}); (\theta, \vec{\pi}))$. The results are dramatically different. For relatively homogeneous courts ($H \leq 0.5$), aggregate welfare at the efficient equilibrium without deliberation for a moderate competence level, $q = 0.80$, actually exceeds aggregate welfare at the most efficient equilibrium with deliberation that is consistent with the data at $q = 0.90$. As the plot shows, the change in welfare is more severe at higher levels of competence. In fact, at $q = 0.9$, the loss of welfare due to deliberation is larger than the change in aggregate welfare that would result from increasing competence from $q = 0.70$ to $q = 0.90$ in equilibria of voting with deliberation consistent with the data.

A similar analysis can be done for $\rho = 0.2$. The previous results show, however, that our previous conclusions do not change when we consider efficient deliberation.



All Equilibria

Equilibria Cons. w/Data

Figure E.8: Maximum Equilibrium Welfare under Deliberation and No-Deliberation. Solid lines denote outcomes with deliberation, for $\rho = 0.5$. Dashed lines denote outcomes with no deliberation.

F Refining the Identified Set: Endogenous Quality of Information.

In Section 5.5 of the main paper we assumed that CASELOAD affects deliberation but is independent of the ability q and preference parameters $\vec{\pi}$. This allowed us to refine the estimated identified set by "intersecting" the estimated identified sets for different levels of CASELOAD. CASELOAD could, however, potentially affect judges' signal quality q. If this were the case, our previous procedure would potentially be too strict and eliminate too many values of $\vec{\pi}$. To reflect this fact, here we consider the possibility that judges' signal quality may depend on CASELOAD, and modify our refinement taking this into consideration. As we will show, provided that preferences $\vec{\pi}$ do not depend on CASELOAD, the revised version of our procedure still provides a significant refinement of the identified set (IS).

We consider two different exercises:

Exercise 1. First, as long as CASELOAD does not affect the preference parameters $\vec{\pi}$, then it still makes sense to take intersections over the identified sets of $\vec{\pi}$ across different values of CASELOAD (i.e., ignore the q parameter). We plot the results of this exercise in Figure F.9.

As the figure shows, the intersection of the identified sets attains a significant reduction in the size of the identified set for (π_1, π_2, π_3), from 4001 to 1403 points. The refinement allows us to rule out the most extreme preferences in favor of overturning, increasing the minimum value of the *bias* of each judge (the smallest value of the posterior probability that the lower court's ruling is erroneous for which judge i would prefer to overturn) from a value of 0.20 for all judges $i = 1, 2, 3$ to 0.40 for judge 1, 0.35 for judge 2, and 0.50 for judge 3. This also implies that we can rule out the most polarized courts in the original identified set. In fact, the highest level of polarization consistent with the data goes from 1.06 (e.g. $\vec{\pi} = (0.92, 0.50, 0.08)$) in the original EIS to 0.67 (e.g., $\vec{\pi} = (0.83, 0.50, 0.17)$) in the refined set.

Strikingly, when we plot the projection on $\vec{\pi}$ of the identified sets (q-slices of which are depicted in Figure A.5) obtained in Section 5.5, where we restrict q to also be unaffected by CASELOAD, the plots in Figure F.9 are exactly the ones that we get. This implies that there are no difference in the identified sets of preference parameters regardless of whether we allow for CASELOAD to affect q. The reason for this is that the identified set for $\vec{\pi}$ is

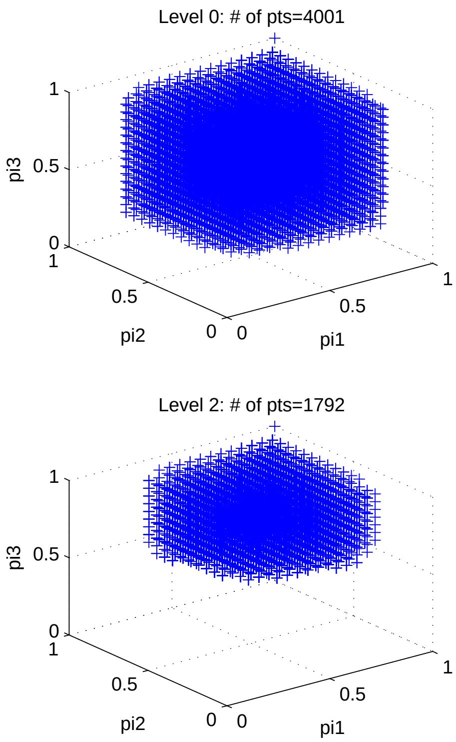


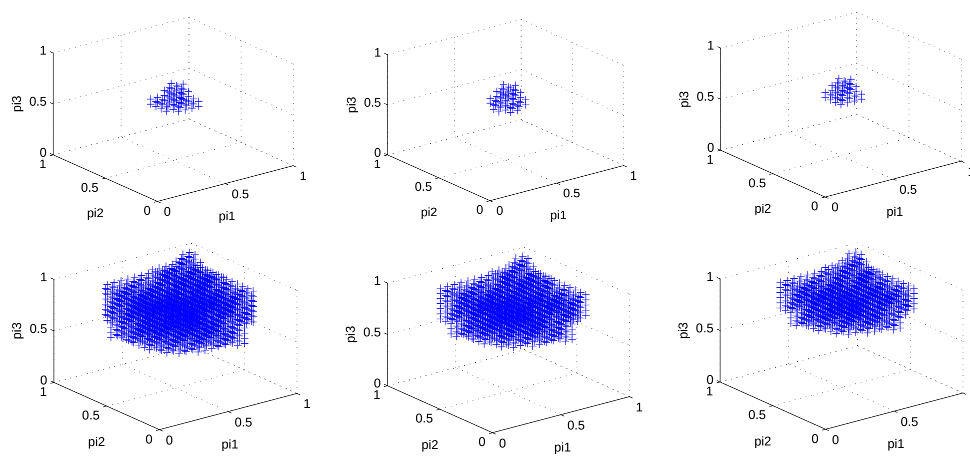
Figure F.9: Level 0 depicts the original EIS of Preference Parameters $\vec{\pi}$. Level 1 depicts the EIS of preference parameters based on median caseload value, level 2 the intersection of median and 25 and 75 quantiles, and level 3 the EIS of $\vec{\pi}$ based on seven quantiles (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875) of the CASELOAD variable in the data.

determined at high values of q, and these high values of q are consistent with the conditional vote profile distribution for all values of CASELOAD.

Exercise 2. For values of q which belong to the identified sets for different CASELOAD levels, our previous procedure remains valid. That is, even when q depends on CASELOAD: the fact that $q \in IS(C')$ and $q \in IS(C'')$ implies that we cannot rule out the possibility that two judges having caseload values of C' and C'' nevertheless have the same value of q. In this case, then, we can use intersection to refine the set of $\vec{\pi}$ consistent with the data *and* the given competence level q. Specifically, define the "q-slice" of the identified set for a value of caseload C, as $IS(C;q) = \{\vec{\pi} \in [0,1]^3 : (\vec{\pi}, q) \in IS(C)\}$; i.e, the set of preference parameters that are consistent with the data for the given level of quality q. Under the assumption that $\vec{\pi}$ does not depend on CASELOAD, if a value of q is in the identified sets for two different values of CASELOAD, the q-slice of the refined set, $\hat{IS}(q)$, only contains values $\vec{\pi}$ that are in both sets; i.e, if $IS(C';q) \neq \emptyset$ and $IS(C'';q) \neq \emptyset$, then $\vec{\pi} \in \hat{IS}(q) \Rightarrow \vec{\pi} \in IS(C';q) \bigcap IS(C'';q)$.

q=0.60

q=0.90



L1: 23284 points   L2: 18296 points   L3: 14048 points

Figure F.10: q-slices of the refined set, $\hat{IS}(q)$, for $q = 0.60$ and $q = 0.90$, with successive refinements as we move from the first to the second and third columns. First column: median; second column: intersection of median and 25 and 75 quantiles; third column: identified set based on seven quantiles (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875) of the CASELOAD variable in the data.

Figure F.10 plots the resulting q-slices of the refined set, $\hat{IS}(q)$ for $q = 0.60$ and $q = 0.90$, with successive refinements as we move from the first to the second and third columns: IS based on median CASELOAD level, IS based on intersection of median and 25 and 75 quantiles of CASELOAD, and identified set based on seven quantiles (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875) of the CASELOAD variable in the data.[4] This allows us to compute a refined set of polarization consistent with the data for each level of q. The result is illustrated in Figure F.11.

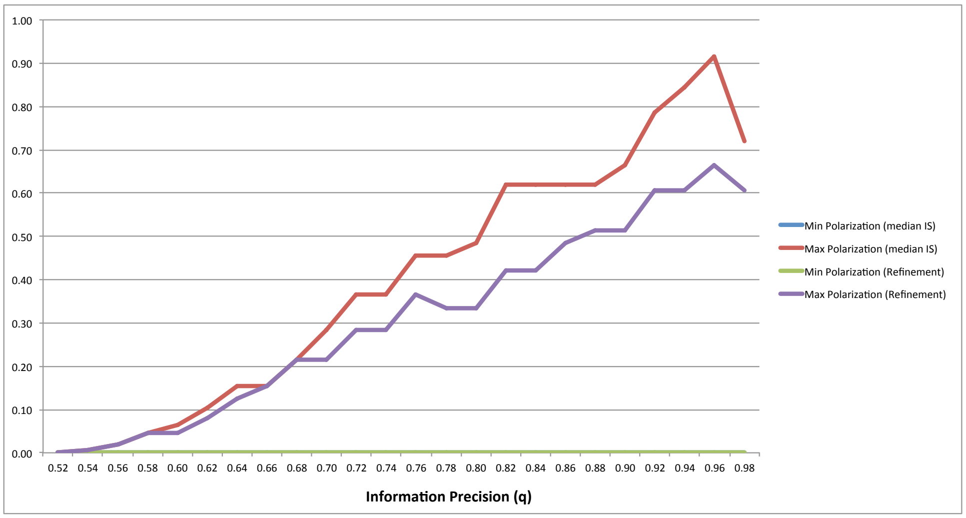


Figure F.11: Min and Max Polarization across points in the median IS and Refined IS

A natural question to ask is whether the effect of deliberation we estimate changes once we use the refined set. The answer is no: qualitatively, the comparison of equilibrium errors with and without deliberation follow similar patterns to what we described in our main analysis (Section 5.4). This is illustrated in Figure F.12 below for two levels of competence ($q = 0.80, 0.90$) and polarization ($H = 0.10, 0.30$)

[4] In practical terms, the results from this second exercise turn out to be very similar to the (unreported) $q = 0.60$ and $q = 0.90$ counterparts of the last three plots of Figure A.5. There is one subtle difference. Previously we assumed exogeneity of q, and computed identified set based on the average vote profile distribution (over all values of caseloads). The average made sense because q does not vary with caseloads. Now that we allow q to depend on caseload, only the conditional vote profile distribution given CASELOAD make sense and only those are used.

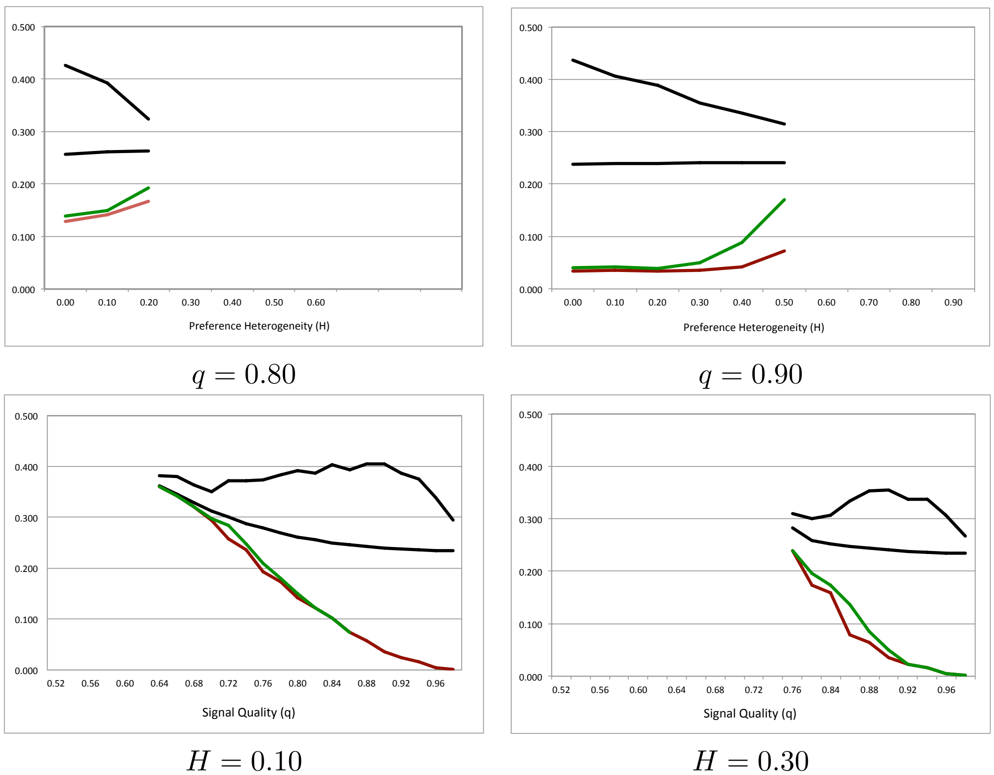


Figure F.12: Eq. probability of mistakes with and without deliberation, for points in the refined EIS, at values of preference heterogeneity H and quality of information q consistent with the refined EIS. $\rho = 0.5$. Min. and max. eq. probability of error (i) in equilibria with deliberation consistent with the data (black), and (iii) in responsive equilibria without deliberation (red and green).

G Common Values Revisited: Results

To address concerns regarding the applicability of our assumptions to some cases in the sample, in particular "admissibility" cases, we re did our estimation and analysis focusing on cases where the common value assumption is most natural according to the discussion in Section 5.5. In this section we report the results of this exercise.

We re-estimate our model (i) for a sample composed of cases other than admissibility cases, and (ii) for a sample of cases including only sufficiency and sentencing (as well as cases coded as other). The rationale for considering the restricted samples is to make sure that the "questionable" cases do not contaminate the estimation of the conditional vote profile distribution for the type of cases that we now focus on. In the new estimates, we compute predicted voting probabilities $p_v(\vec{v}|X)$ for sufficiency cases.

Table G.6 below presents the results. Column (1) presents the predicted voting probabilities for sufficiency cases using the full sample. Columns (2) and (3) show the corresponding estimates for our two robustness checks. For ease of comparison, the predicted voting probabilities for our benchmark analysis is also reported in Column (4). The comparison of columns (1) and (4) show that predicted vote probabilities for sufficiency cases differ somewhat from those for admissibility cases, although not radically. Moreover, comparing column (1) with columns (2) and (3) shows that excluding questionable cases similarly does not have a large affect on predicted vote probabilities.

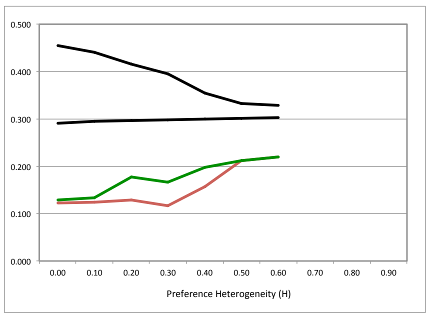
Table G.6: Estimated Vote Probabilities $p_v(\vec{v}|X)$ using different subsamples

Subsamples of Cases:	(1) **Full Sample**	(2) **Only Sufficiency & Sentencing**	(3) **All Except Admissibility**	(4) **Full Sample**
$\hat{p}_v(111)$	0.194	0.189	0.169	0.223
$\hat{p}_v(101)$	0.010	0.016	0.014	0.020
$\hat{p}_v(110)$	0.007	0.007	0.008	0.013
$\hat{p}_v(100)$	0.014	0.027	0.023	0.025
$\hat{p}_v(000)$	0.746	0.707	0.743	0.677
$\hat{p}_v(010)$	0.008	0.016	0.012	0.015
$\hat{p}_v(001)$	0.011	0.021	0.015	0.018
$\hat{p}_v(011)$	0.013	0.017	0.016	0.010

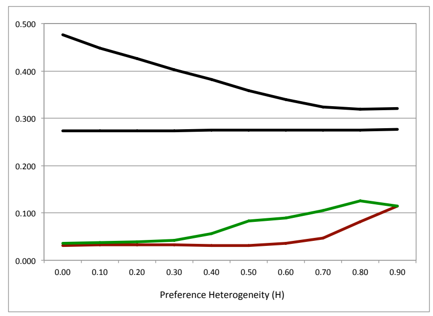
Columns (1)-(3): Probabilities evaluated at case characteristics
$X =$ (FedLaw, White Collar, Sufficiency, Rep. Majority)

Column (4): Probabilities evaluated at case characteristics
$X =$ (FedLaw, White Collar, Admissibility, Rep. Majority)

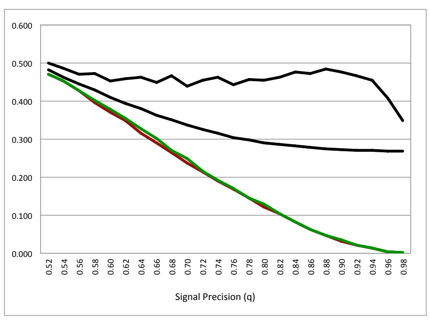
The moderate change in the voting probabilities strongly suggests in turn that the changes in our estimates of the effect of deliberation will also be moderate, but does not imply this result. To confirm this, we recompute the max and min error probabilities in equilibria with and without deliberation for the new estimated voting probabilities. We choose the second column of Table G.6 to redo the error probability calculations with, as this column comes from a configuration that ex ante differs from the benchmark configuration the most. The results, which we illustrate in Figure G, indicate that our main conclusions are qualitatively unchanged.



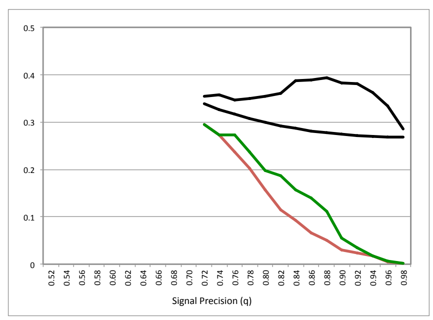
$q = 0.80$



$q = 0.90$



$H = 0$



$H = 0.4$

Figure G.13: Probability of mistakes with and without deliberation for values of preference heterogeneity H and quality of information q consistent with points $(\vec{\pi}, \theta)$ in the EIS for the sufficiency cases based on the most restricted sample. $\rho = 0.5$. Min. and max. eq. probability of error (i) in equilibria with deliberation consistent with the data (black), and (iii) in responsive equilibria without deliberation (red and green).

References

Bugni, F., I. Canay, and X. Shi (2013): "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," Working Paper, Duke University.

Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kitamura, Y., and J. Stoye (2011): "Nonparametric Analysis of Random Utility Model," Discussion paper, Cornell University.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Applications," *Econometrica*, 83(1), 315–334.

Shi, X., and M. Shum (2015): "Simple Two-Stage Inference for A Class of Partially Identied Models," *Econometric Theory*, 31, 493–520.

Wolak, F. (1989): "Testing Inequality Constraints in Linear Econometric Models," *Journal of Econometrics*, 41, 205–235.