Should Political Scientists Use the Self-Confirming Equilibrium Concept? Explaining the Choices of Cognitively Limited Actors

Lupia, Arthur and Zharinova, Natasha and Levine, Adam Seth

University of Michigan

28 January 2007

Should Political Scientists Use the Self-Confirming Equilibrium Concept? Explaining the Choices of Cognitively Limited Actors

Arthur Lupia, Natasha Zharinova, and Adam Seth Levine

Biographical Information

The corresponding author is Arthur Lupia.

Arthur Lupia is the Hal R. Varian Collegiate Professor of Political Science at the University of Michigan. Address: 4252 Institute for Social Research, 426 Thompson Street, Ann Arbor, MI 48104-2321 (lupia@umich.edu).

Natasha Zharinova is an advisor at the Risk Advisory Services at ABN AMRO Bank. Address: RAS, ABN AMRO Bank N.V., Gustav Mahlerlaan 10, P.O. Box 283, 1000 EA Amsterdam, The Netherlands (natalia.zharinova@nl.abnamro.com).

Adam Seth Levine is a graduate student in the Department of Political Science, University of Michigan. Address: 4252 Institute for Social Research, 426 Thompson Street, Ann Arbor, MI 48104-2321 (adamseth@umich.edu).

Abstract

Many claims about political behavior are based on implicit assumptions about human reasoning. One such assumption, that political actors think in complex and similar ways when assessing strategies, is nested within widely used game theoretic equilibrium concepts. Empirical research casts doubt on the validity of these assumptions in important cases. For example, the finding that some citizens expend limited cognitive energy to social concerns runs counter to the assumption that citizens (e.g., jurors) base all decisions on complex thoughts. Similarly, evidence that some political actors (e.g., Democrats and Republicans) think about political cause-and-effect quite differently runs against the assumption that all players reason about politics in similar ways. The self-confirming equilibrium (SCE) concept provides a means for evaluating the robustness of theoretical conclusions to the introduction of a broad range of psychological assumptions. Through arguments and examples, we explain opportunities and challenges inherent in using the SCE concept. We find that the concept provides an improved foundation for more serious and constructive interactions between formal theoretic and psychology-oriented literatures.

In many game theoretic studies of politics, the choice of an equilibrium concept is equivalent to making a special set of assumptions about how people think. Many theorists adopt the Nash Equilibrium concept, which can entail the assumption that all players think in a very rigorous manner (see, e.g., Turner 2000, 2001). This reasoning involves all players in a game basing their strategy not only on extensive knowledge of the game itself but also on complex and nearly-identical conjectures about what all other players in will do (Aumann and Brandenberger 1995).

How many political actors think in such ways? Clearly, some do not. Citizens who have little interest in politics do not appear to base their turnout or voting decisions on detailed assessments of what other political actors will do. It is also clear that citizens can base their political decisions on distinctly non-identical beliefs about nature and/or conjectures about the behaviors of others. Religious conservatives and humanist liberals, Israelis and Palestinians, and rich and poor are among pairs of politically relevant groups who think about important aspects of politics in very different ways.

How should these facts affect game-theoretic political science? It depends. We agree with those who argue that many people do not literally engage in the kind of reasoning that common equilibrium concepts presuppose (see, e.g., Rubinstein 1998). We also agree with those who claim that some political actors make decisions "as if" such reasoning occurs. We agree, for example, with Satz and Ferejohn (1994) who argue that institutions can structure choices in a way that give people an incentive to think about their options in ways that are consistent with Nash-based assertions. In some cases, however, the "as if" claim is hard to justify. It is possible that all citizens who pay limited attention to politics reason “as if” they think in complex terms (e.g., simple “rules of thumb” yield equivalent outcomes). But what if some do not?

A generation of theorists has recognized such challenges and taken steps to meet them. Some, like Harsanyi (1967, 1968) and Kreps and Wilson (1982), have refined the Nash concept to allow players to choose best responses even though they lack information about specific aspects of the game. Others, such as Aumann (1974) and McKelvey and Palfrey (1995), have diverged farther from the basic Nash concept. Given the frequency with which political scientists encounter actors who may base their decisions on little or no cognitive effort -- as well as actors who share a decision context despite having polar-opposite worldviews – it is reasonable to question whether commonly-used equilibrium concepts provide the most effective means for characterizing all kinds of political behavior.

We argue that political science can benefit by turning some of its theoretical energies to an alternate approach. One such approach entails using the self-confirming equilibrium concept (Fudenberg and Levine 1993, 1998, Dekel, Fudenberg, and Levine 1999, 2004; henceforth, SCE). The key element of a SCE is the correspondence between what a player does and what she observes. If her observations are consistent with her beliefs about nature and her conjectures about the actions of other players, then her rationale for her actions is positively reinforced. If *all* players receive such reinforcement, then their actions are “in equilibrium.”

Like Nash-based equilibrium concepts, a SCE characterizes players as strategic – in the sense that they are depicted as basing a plan of action on what they believe, conjecture, and observe.[1] The difference is that we can use a SCE to explain the choice of players who know very little about one another and who need not think rigorously about what little they know. In a SCE, players are not presumed to think of all other players in complex or similar ways.

We find that researchers of politics can benefit by considering equilibrium concepts that allow a wider range of reasoning mechanics. In this sense, our work echoes previous arguments by political methodologists. To see how, recall that at one time most multivariate estimations in our discipline were conducted using ordinary least squares (OLS). Then, scholars imported from other disciplines the critique that some kinds of data were incompatible with the method (i.e., the data's variance structure violated one or more of the assumptions that validates OLS estimates). These discoveries did not invalidate OLS generally, but they caused our discipline to seek more appropriate estimators for the class of problems in question. For example, it is now common practice for scholars to use maximum likelihood estimators when their dependent variable is limited --- e.g., binary -- and discrete. Here, we argue that key assumptions about reasoning on which the validity of common equilibrium concepts are based are violated in some political contexts. In such cases, SCE can be to game theorists what logit and probit are to empirical scholars – a more effective and credible means for drawing reliable conclusions.

We proceed as follows. First, we describe reasoning assumptions that are implicit in common equilibrium concepts. Second, we present the SCE concept. In the process, we offer examples where basing inferences on the SCE concept leads to different, but

[1] This attribute distinguishes use of the SCE concept from other variants of the Nash Equilibrium idea such as rationalizability (Pearce 1984 and Bernheim 1984).

constructive, insights about important political questions. In each of our examples, the findings are more than a technical curiosity – they come from attempts to reconcile a formal model with empirically defensible assumptions about how political actors think.

In the main example, we use SCE to cultivate a link between psychological and game-theoretic studies of jury decision making. We offer a variant of a recent jury model (Feddersen and Pesendorfer 1998). The source of our variant is psychological research on how jurors process trial information (e.g., Pennington and Hastie 1993) and on differences in how deeply citizens think in such situations (Cacioppo and Petty 1982). We then use a SCE to characterize behavior and outcomes in a formal model that includes the kinds of jurors that these scholars have observed. As a result, this example provides a framework for integrating the game theoretic and psychological literatures, while clarifying whether theoretical concerns about the normative virtues of unanimous verdicts are robust to the presence of the kinds of jurors observed by psychologists.

In sum, game theory generates important insights about many political phenomena. These insights are judged by many criteria. A criterion in which there is growing interest is psychological viability (see, e.g., Camerer, Loewenstein, and Rabin 2004). With this criterion in mind, we can agree that the credibility of game-theoretic political science need not rest on the sometimes-untenable assumptions about human reasoning that are embedded in common equilibrium concepts. For cases where political actors do not reason as these concepts presume, there is value in exploring alternate concepts whose bases are more easily reconciled with the underlying psychology.

Cognitive Foundations of Nash-Based Concepts

For many people, game theory and the Nash Equilibrium concept (henceforth, NE) are synonymous. Given the frequency with which the concept is used in game theoretic political science, the perceived synonymy is justifiable. NE, however, is just one of several often-used equilibrium concepts.

While many non-cooperative game theoretic studies in political science do not use NE, almost all use refinements of the Nash concept. Common refinements include the subgame-perfect, trembling-hand perfect, Bayesian-Nash, perfect Bayesian, and sequential equilibrium concepts. Subgame perfection, for example, is a NE refinement that strengthens the inferential power of game theoretic treatments in extensive form games – where strengthening implies introducing an additional technical criterion that is appropriate for that class of games. The other attribute of these refinements is that they retain core properties of the original NE concept – in particular, its requirement that player strategies constitute best responses to the strategies of all other players – with the response evaluated *along the equilibrium path* in games containing sequences of moves.[2]

Why do so many theorists choose Nash-based equilibrium concepts? Two explanations are focal. First, these concepts induce the theorist to focus on identifying *steady states* in a game. Steady states make theoretical conclusions more persuasive. After all, if showing players your prediction of their behavior in a game causes one or more of them to change their behaviors, then the prediction is not very reliable (i.e., the state is not steady). Nash-based equilibrium concepts preclude such occurrences. Second, Nash himself (1950) proved that Nash equilibria exist in any finite extensive form game.

[2] For simplicity, we use the term "equilibrium path" to characterize paths of any length (including zero) which allows us to use a single term to cover equilibria in all normal and extensive form games.

For these and other reasons, many people treat Nash-based concepts as substantively innocuous -- as entailing no substantive baggage. This is wrong. Each of these concepts presumes that players reason in a specific manner. To see how, consider Gibbons' (1992: 8-9) definition of a Nash Equilibrium, where S_i denotes the set of possible strategies for player *i*, s_i denotes an element of that set, and $u_i(s_1,..s_n)$ denotes player *i*'s utility function and refers to the fact that her utility can be a function of other player's strategies as well as her own.

> "In the n-player normal-form game $G=\{S_1,...S_n; u_1,...u_n\}$, the strategies $(s_1*,...s_n*)$ are a Nash equilibrium if, for each player *i*, s_i* is (at least tied for) player *i*'s best response to the strategies specified for the *n-1* other players, $(s_1*,...s_{i-1}*, s_{i+1}*,...s_n*)$: $u_i(s_1*,...s_{i-1}*,s_i*, s_{i+1}*,...s_n*)\geq ui(s_1*,...s_{i-1}*, s_i, s_{i+1}*,...s_n*)$ for every feasible strategy s_i in S_i; that is, s_i* solves $max_{s_i \in S_i}\ ui(s_1*,...s_{i-1}*, s_i, s_{i+1}*,...s_n*)$."

Here, each player's strategy is a function of every other player's strategy along the equilibrium path and each strategy is a complete plan of action that details what the player will do at possibly many other points in the game. In cognitive terms, the Nash criterion requires that all players base their decisions on a potentially rigorous conjecture. We say rigorous because they entail all players imagining all other players' decisions at various points in a game. In games that have many players, or games in which even a few players can take many actions, these conjectures can be quite complex.

Many theoretical claims are based on more than a presumption of rigorous conjectures. They also require *shared conjectures.* As Aumann and Brandenberger (1995: 1163) describe,

> "In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures [about the strategies of others] are commonly known. Then for each player *j*, all the other players *i* agree on the same conjecture σ_j about *j*; and the resulting profile $(\sigma_{1,...,}\ \sigma_n)$ of mixed actions is a Nash equilibrium."

So, for a set of strategies to be in equilibrium, not only must every player have and share a potentially complex mental model of what every other player will do at other decision nodes, but it is also the case two players cannot disagree about the actions of a third along the equilibrium path.

Other Nash refinements have similar attributes. While the main way in which common refinements (e.g., trembling-hand perfection, Bayesian-Nash and sequential equilibria) differ from one another is in what they allow players to conjecture about others, these refinements continue to posit that actors run shared and often complex cognitive assessments. Consider, for example, the Bayesian Nash Equilibrium (BNE), which is used in games of incomplete information. To simplify the comparison of the concepts, we again return to Gibbons' (1992: 151) notation. Here, in addition to the notation described in his NE definition, A_i denotes the set of possible actions for player *i*, a_i denotes an element of that set, Θ_i denotes the set of possible types for player *i*, θ_i denotes an element of that set, and p_i denotes the common knowledge prior belief that player *i* has a particular type (i.e., p_i is a probability distribution over Θ_i). A "type" refers to a potentially game-relevant attribute (e.g., a policy preference or level of knowledge) or set of attributes of a player about which other players can be uncertain. Let Θ_{-i} denote the set of possible types for all players except *i*, where θ_{-i} denotes an element of that set. In a game of incomplete information, at least one player is uncertain about another player's type (another player's attributes) or Nature's type (contextual attributes).

> In the static Bayesian game $G=\{A_1,\ldots,A_n;\ \theta_1,\ldots\ \theta_n;\ p_1,\ldots,p_n;u_1,\ldots,u_n\}$, the strategies $s^*=(s^*_1,\ldots,s^*_n)$ are a pure strategy Bayesian-Nash equilibrium if for each player *i* and for each of *i*'s types $\theta_i \in \Theta_i$, $s^*_i(\theta_i)$ solves $max_{ai \in Ai}\ \Sigma_{\theta_{-i} \in \Theta_{-i}}$ $u_i(s^*_1(\theta_1),\ldots,s_{i-1}(\theta_{i-1}),a_i,s^*_{i+1}(\theta_{i+1}),\ldots s^*_n(\theta_n);t)p_i(\theta_{-i}|\theta_i)$.

A steady state occurs when no player has an incentive to change his or her strategy, *even if the change involves only one action by one possible type of one player*. Here, each player's choice depends on a conjecture about what *every possible type* of every player will choose. If a player has a thousand possible types, then all other players must base their strategies on a conjecture of the play of each of that player's thousand possible types. While they need not know which type of a particular player they are actually playing against, the equilibrium concept presumes their decision to be the one that would result if they had thought through what every other person would do in equilibrium given every possible type that every other person could have.

Players in a BNE must also share conjectures. Not only can two players not disagree about the posited equilibrium play of a third (or Nature), they cannot disagree about the posited play of even one of the third player's (or Nature's) possibly infinite number of types. *All players must share conjectures about what every single type of every single player would do at every decision node* that is relevant to the posited equilibrium. In sum, such assessments can entail quite a calculation!

Given what has been written about the apathy of common citizens, it is reasonable to ask how many base their actions on rigorous and shared beliefs and conjectures of the kinds described above. Reasoning requires time, effort and at least a modicum of cognitive energy. Even for motivated people, information processing is characterized by severe constraints (see, e.g., Kandel, et. al. 1995: 651-666). Chief among these constraints are the very limited storage capacity and very high decay rates of working memory as well as the rules by which certain stimuli gain access to long-term memory.[3]

[3] Bjork and Bjork (1996) and Schacter (1996, 2002) provide entry-level references for properties of memory and their implications for social interaction.

One implication of these attributes is that political ideologues are likely to pay attention to different stimuli, remember different events, and create and reinforce different internal theories of cause-and-effect (Pennington and Hastie 1990, 1993).

To be sure, some political actors reason "as if" they are rigorously thinking through the possible strategies of all other players. Others likely process information in ways that yield identical conjectures about what everyone else is doing. Just as surely, other actors think differently. In such cases, a different approach can be informative.

The SCE concept is such an approach. It yields steady state conclusions about social behaviors without requiring traditional reasoning assumptions. When empirical evidence suggests that political actors spend limited effort processing complex stimuli – or process such information in different ways – SCE can offer a more credible foundation for explaining their behavior.

A Definition and Implications of the Self-Confirming Equilibrium Concept

Fudenberg and Levine (1993, 1998), Fudenberg and Kreps (1995), and Dekel, Fudenberg and Levine (1999, 2004) developed the self-confirming equilibrium concept in a series of papers and books whose publication dates span a decade. In this section, we offer a brief primer on the concept. Our main reference is Dekel, Fudenberg, and Levine (2004), which we denote as DFL.

Let *i* be a player in the game and let *I* be the set of players in the game, where the number of players is finite. Let $\theta_i \in \Theta_i$ be player *i*'s type, let $\theta_0 \in \Theta_0$ be Nature's type, and let θ_{-i} be the vector of other players' types. Let $a_i \in A_i$ denote player *i*'s action and let $\sigma_i(a_i) \in \Sigma_I$, henceforth σ_i, denote a mixed strategy for player *i* in the set of possible mixed strategies for her.

The game's common knowledge includes players knowing their own utility functions. *The common knowledge need not include much else.* It need not include the strategies used by other players or other exogenous contextual attributes (a.k.a., moves by Nature). Specifically, let $\mu_i \in \Delta(\Theta_0)$ be player *i*'s prior belief about Nature's move and let *r* be the true state of Nature. If $\forall i,j \in I,\ \mu_i=\mu_j$, then we say that players have common prior beliefs. When $\mu_i=r$, player *i* has correct prior beliefs about Nature. We need not (and do not always) assume common or correct prior beliefs in what follows.

Let $y_i=y_i(a,\theta)$ be player *i*'s "private signal" about the play of the game. This signal is what player *i* observes in the game. This signal can include any or all of the following: which terminal node is reached, information about previous moves or previous plays, a player's own payoffs, and other players' payoffs. It may also include none of the above – an assumption we can make if we want to model a situation where a player either receives no feedback about a game or pays no attention to the feedback that is available to him. The term "private signal" when used in a SCE context is <u>not</u> equivalent to the term "private information" that is often used to describe aspects of a game that are known to one player but not another. While the information contained in a private signal can be private information, it need not be – it can be known by others.

Finally, let $\hat{\sigma}_{-i} \in \times_{-i}\Sigma_{-i}$ be player *i*'s conjecture about his opponents' play (specifically, his conjecture about the strategy profile of his opponents) and let $u_i(a_i,\theta)$ be player *i*'s expected utility from playing a_i. We now have sufficient definitions and notation to present DFL's (p. 286) definition of a self-confirming equilibrium.[4]

[4] We restrict attention to what DFL (p. 287) call SCE with independent beliefs, which implies that player i's beliefs about her opponents' types do not depend on her own type. This independence restriction parallels an assumption made in nearly all games of incomplete information in political science.

Definition: A strategy profile σ is a *self-confirming equilibrium with conjectures* $\hat{\sigma}_{-i}$ *and beliefs* $\hat{\mu}_i$ if for each player *i,*

(i) $\forall \theta_i, r(\theta_i) = \hat{\mu}_i(\theta_i)$, *and* for any pair $\theta_i, \hat{a}_i$ such that $\hat{\mu}_i(\theta_i) \cdot \sigma_i(\hat{a}_i \mid \theta_i) > 0$ both of the following conditions are satisfied

(ii) $\hat{a}_i \in argmax\ a_i \sum_{a\text{-}i, \theta\text{-}i} u_i(\hat{a}_i, a_{-i}, \theta_i, \theta_{-i})\ \hat{\mu}_i(\theta_{-i} \mid \theta_i)\ \hat{\sigma}_{-i}\ (a_{-i} \mid \theta_{-i})$ and $\forall\ \overline{y}_i$ in the range of y_i

(iii) $\sum \{a_{-i}, \theta_{-i}: y_i(\hat{a}_i, a_{-i}, \theta_i, \theta_{-i}) = \overline{y}_i\}\ \hat{\mu}_i(\theta_{-i} \mid \theta_i)\ \hat{\sigma}_{-i}\ (a_{-i} \mid \theta_{-i}) = \sum \{a_{-i}, \theta_{-i}: y_i(\hat{a}_i, a_{-i}, \theta_i, \theta_{-i}) = \overline{y}_i\}\ r(\theta_{-i} \mid \theta_i) \sigma_i\ (a_{-i} \mid \theta_{-i})$

In words, a SCE has three requirements. *Condition i* states that each player has correct beliefs about her own type. *Condition ii* states that any action that a player plays with positive probability must maximize her utility given her conjecture and beliefs. *Condition iii* (hereafter *C3*) describes qualities of player conjectures that are allowable in equilibrium. *C3* is the key difference between SCE and common Nash refinements.

While *Condition ii* requires that each player's strategy be a best response to the player's conjectures about opponents' play, *C3* requires these conjectures be consistent with the player's observations. When a player's observations, beliefs, and conjectures are in synch, what she sees confirms her choice and gives her no reason to change -- and what a player does never results in her observing something unexpected. When the same is true for all players, then the strategy profile is in equilibrium. In a SCE, therefore, each player's strategy is a best response *to her own beliefs, conjectures, and observations* (if any) and not necessarily to the strategies of other players. To satisfy *C3*, it is sufficient that player conjectures and observations are consistent. How they become consistent – whether through conjectures that are shared, unshared, simple or complex -- is irrelevant.

For game-theoretic political science, a *SCE* has four critical properties: observations must be consistent with the interaction between beliefs and conjectures, incorrect conjectures are allowed, two players can disagree about a third (or Nature), and more precise observations by players imply greater constraints on what constitutes a SCE. We address the substantive implications of each property in turn.

The Relationship between Observations and Conjectures

> "[E]ach player attempts to maximize his own expected utility. How he should go about doing this depends on how he thinks his opponents are playing, and the major issue … is how he should form those expectations" (Fudenberg and Levine 1998:14).

The SCE requires that players' expectations are formed by their beliefs and conjectures and confirmed by their observations. A motivation for this move is as follows:

> "The most natural assumption in many … contexts is that agents observe the terminal nodes (outcomes) that are reached in their own plays of the game, but that agents do not observe the parts of their opponents' strategies that specify how the opponents would have played at information sets that were not reached in that play of the game… [I]n many settings players will not even observe the realized terminal node, as several different terminal nodes many be consistent with their observation" (Fudenberg and Levine 1998:175).

This description clearly applies to goal-oriented political actors who have insufficient data or incentive to deduce other players' actions or types from their observations.

Another way to state this implication is that players in a SCE do not explicitly contemplate each other's rationality. Unlike Nash-based concepts, they need not think about explicit moves by other players at particular decision nodes. They need not justify their strategies as best responses to the anticipated strategies of other players. In a SCE, players just need a theory of cause and effect that keeps them from making mistakes that they can recognize. If an actor has imprecise feedback, then she may choose actions that she would view as sub-optimal if better informed. Nevertheless, if what she sees is consistent with what she believes and conjectures, she has no rationale for changing her strategy -- her behavior can be described in terms of a steady-state relation.

Of course, we can imagine cases where actors would be hesitant to base their conjectures on uninformative private signals. If such actors had opportunities to improve their feedback, they would do it. Fair enough. But many actors that political scientists study choose not to learn about politics nor do they put much effort into thinking about the political information they receive. Alternate equilibrium concepts, such as SCE, can help theorists model such actors more effectively.[5]

Incorrect Beliefs and Conjectures are Allowed

An important difference between the SCE concept and more common equilibrium concepts is that actors in a SCE can maintain incorrect beliefs and conjectures in a steady state. So where common Nash-based treatments require $\hat{\mu}_i = r$ and $\hat{\sigma}_i = \sigma_i$ -- all players must share correct conjectures about the action that every single type of every single player would choose at every decision node along the equilibrium path -- variance in the quality of the private signal allows players to maintain incorrect beliefs and conjectures in equilibrium.

A maximizing strategy in a SCE can include a suboptimal action at information sets that would be reached with probability zero if the actor's conjecture were true. The rationale for maintaining such a strategy in a SCE is that the actor does not expect to learn that her conjecture is untrue -- and if she never does, then she has no reason to change strategy. We can certainly imagine political actors whose cognitive approach to

[5] Most non-cooperative games of incomplete information use refinements of the Nash equilibrium concept (e.g., perfect Bayesian equilibrium, sequential equilibrium) that presupposes use Bayes' Rule to draw inferences about Nature and other players along the equilibrium path. The SCE concept, by contrast, does not require that actors use Bayes' Rule. It requires only that actors' beliefs and conjectures, however drawn, are consistent with their observations. In other words, when Bayesian updating is assumed, posterior beliefs are constrained to have a specific functional relationship to prior beliefs. In a SCE, things are different. To the extent that a player's private signal is generated by reality (i.e., the true distribution of Nature's and/or players' types), it is not correct to say that the SCE outcome must be independent of prior beliefs. However, in a SCE the relationship between priors and posteriors can be far less direct that Bayes' Rule posits.

politics has such attributes. If, for example, *all* of a voter's evidence suggests that their rule-of-thumb (e.g., vote Republican) yields good choices, why should they think any more about it?

To some readers, such a statement may seem to be an anathema. Game theory, after all, is often linked with the idea of rationality. The maintenance of incorrect conjectures and potentially suboptimal strategies will strike some readers as anything but rational. To such reactions, one thing is worth pointing out. A problem with many claims about "rationality" is that there are numerous conflicting definitions of the term in circulation (see, e.g., the definitional inventory in Lupia, McCubbins, and Popkin (2000: 3-11)). Among the least useful of these definitions for explaining the actions of flesh-and-blood human actors are definitions that equate rationality and omniscience. Alternative definitions hold rationality as the product of human reason, where reason is the ordinary function of the mind. Therefore, it is a reader's positing of omniscience as a desirable analytic standard, rather than a search for properties of standard human reason, that makes a game-theoretic steady state that features incorrect conjectures appear inconsistent. One does not have to be omniscient to act strategically. A SCE allows us to characterize actors as goal oriented, strategic, and in possession of recognizable, but imperfect, cognitive endowments.

Two players can disagree about attributes of a third

Unlike common Nash-based concepts, two players in a SCE can disagree about the actions or types of a third. For example, in a three player game where a player's shirt color affects player payoffs, Player 1 can believe that Player 3 is wearing a blue shirt, Player 2 can believe that Player 3 is wearing a yellow shirt, and as long as Player 1 and

2’s observations are consistent with these beliefs (which means that private signals could not include player 3’s shirt color), neither player has an incentive to change their actions or beliefs. To see why this factor matters, consider a simple example that shows the impact of moving from Nash-based equilibrium concepts to SCE. In Figure 1, Congress and the President are in a standoff over the budget. If the standoff persists, as it did in the mid 1990’s, the government will shut down, which hurts many voters.

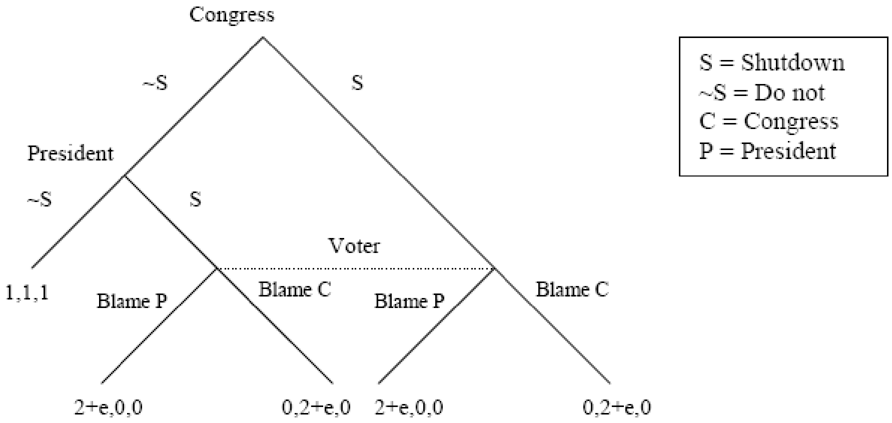

Figure 1. Congress-President Standoff

If Congress and the President end the standoff, then all players earn a payoff of 1. If either player continues the standoff, the government shuts down and the move goes to a representative voter. The voter, who observes a government shutdown, but not why it occurred, blames either the President or the Congress. The player who is not blamed benefits with a payoff of *2+e*, where *e>0* and can be very small.

The outcome (~s, ~s) (i.e., Congress and the President agree to end the standoff) is a SCE when Congress conjectures that it is more likely than the president to be blamed for the standoff while the President conjectures that he is more likely than Congress to be blamed. Since the voter’s decision node is not reached in equilibrium, everything the

Congress and the President observe is consistent with their conjectures (i.e., they never observe the other being blamed). Their choices of strategy are confirmed.

The outcome, (~s, ~s), could not occur in models where the equilibrium concept required Congress and the President to have identical conjectures about the voter's move. To see why, note that the standoff continues if the voter blames either player with probability greater that or equal to probability 1/2, which the voter cannot help but do since the two probabilities (the probability of blaming Congress and the probability of blaming the President) must sum to 1. Therefore, any mixed strategy by the voter would induce at least one of the other two players to continue the standoff.

It is worth noting that if *e* is sufficiently small, then producing the result above requires only a small difference between presidential and congressional conjectures about voter behavior. Each entity could, for example, conjecture that the likelihood of its being blamed was 51%. This 2% difference is within the margin of error of even the best political polls and, as such, can be smaller than the difference in polls that each entity might commission in reality.

Not only can each player maintain different conjectures about a third in a SCE, but *different types of the same player* can maintain varying conjectures about Nature or other players. For politics, this aspect of the SCE permits greater flexibility in representing the mindsets of different types of people who can inhabit the same player roles – such as that of a pivotal voter. In such roles, we can imagine lifelong Democrats and lifelong Republicans basing their strategies on very different notions of political cause and effect (e.g., why George W. Bush pursued a war with Iraq). SCE allows us to derive steady state characterizations of players who share neither common prior beliefs nor identical reasoning algorithms.

As the precision and range of observation decreases, so do SCE constraints

The correspondence between a game's NE and SCE depends on what players observe. In general, the more players observe, the closer is the correspondence. The theoretical implications of this correspondence become clear in extreme cases.

At one extreme, suppose that the private signal is completely informative. In this case, NE and SCE coincide. That is, when the play of the game reveals a player's own payoff and other players' strategies, then utility maximization implies choosing a best response to other players' (observed) strategies. Moreover, when players' private signals fully reveal Nature's move in a game where at least one player (or Nature) has multiple types, then SCE and BNE are equivalent as well. Here, the actor must maximize utility with respect to every player type that they expect to encounter. In general, the more they know, the more their learning must resemble Bayes' Rule in equilibrium.

As the private signal becomes less informative, NE and SCE often diverge. At the extreme, when private signals are uninformative, the set of SCE allows all profiles of *ex ante* undominated strategies (DFL: Proposition 1). In other words, players attempt to maximize utility, but without the feedback we normally think of game-theoretic actors possessing. Put another way, SCE is not a NE refinement.

In many games, the SCE yields a larger set of equilibria than do equivalent models characterized using better-known equilibrium concepts. In fact, the set of SCE is often a superset of the set of NE. Since game theorists often view expansions of the set of equilibria as a bad thing – as is evidenced by the amount of effort spent on equilibrium refinement – it is reasonable to ask whether the extra strategy profiles that emerge merit scholarly attention.

The answer to this question depends on how much one values deriving theoretical conclusions from empirically-defensible premises. When the set of SCE and NE of the same game differ, the difference is a result of loosening the Nash-based concept's reasoning requirements. When empirical evidence or other theory demonstrates that the people whose behavior a model is constructed to explain do not reason "as if" they share rigorous Nash-style conjectures, the change in the set of equilibria caused by moving to SCE is a signal that the smaller set of *Nash-based conclusions were artifacts of unrealistic cognitive assumptions*. So, ignoring the SCE's conclusions in such cases is akin to sacrificing the argument's soundness. Put constructively, it is important to pay attention to the set of SCE in such cases because they provide a means for evaluating the robustness of extant Nash-based conclusions to more realistic assumptions about how people think. In our final example, we use the SCE concept for just this purpose.

Example: Thinking Differently in Jury Theorems

In this example, we briefly re-examine an important question about jury decision making. This topic has received great attention from game theorists in recent years. Psychologists have also studied it extensively. The psychological research reveals significant variations in how jurors think. But the theoretical and psychological literatures do not speak to one another. As a result, theoretical consequences of observed variations in how jurors think have not been explored. Our brief SCE-oriented example draws insights from both research traditions in an attempt to clarify these consequences.

Background

The focus of current jury theorems begins with the Condorcet Jury Theorem (1785; henceforth CJT). In it, a jury of n members chooses one of two alternatives, say A or C (i.e., acquit or convict). It is common knowledge that one of these alternatives corresponds to the true state of the world (innocence or guilt) and that everyone prefers the group to choose that alternative. But the true state of the world need not be known. The CJT shows that if the probability of each member choosing the "better alternative" is greater than .5, then the probability that a majority will choose it goes to 1 as $n \Rightarrow \infty$. The result highlights beneficial information aggregation properties of common collective decision rules.

Austen-Smith and Banks (1996) showed that information aggregation need not be so beneficial. Their analysis begins with a question about whether individuals make the same choices when voting as a member of a jury as they do when voting alone. Austen-Smith and Banks then model each juror as receiving an evidentiary signal, say $m_j \in \{G,I\}$, that conveys information about the true state of the world (i.e., guilty or innocent).[6] Substantively, the signal represents a juror's view of trial evidence and deliberation. Technically, each juror's signal is determined by a single, independent draw from a Bernoulli distribution. While it is assumed that each juror observes only their own signal, two things about the distribution are commonly known. First, the true state of the world is G with probability $s \in (0,1)$ -- and is I with probability $1-s$. Second, each signal gives the true state of the world with probability $p \in (.5,1)$ – and the false one with probability $1-p$.

Austen-Smith and Banks' work investigates whether all jurors in this circumstance would vote to convict when $m_j=G$ and vote to acquit when $m_j=I$. If all jurors

[6] We use the term "evidentiary signal" to describe what the jury models call a "private signal" to avoid confusion with the SCE literature's long-standing, but distinct, use of the same term.

were to vote in accordance with their evidentiary signals, the CJT's beneficial information aggregation properties would survive. But Austen-Smith and Banks show that such behavior *need not* be a Nash equilibrium. Their finding comes from seeing a juror as being in one of two situations: "pivotal" or "not pivotal." If a juror is "not pivotal," then her vote cannot affect the verdict and what she does with her information has no bearing on whether or not the group chooses the better alternative. By contrast, if the juror is "pivotal" and majority rule is being used, then the aggregate outcome is a tie without her vote. In this case, if everyone else is voting in accordance with their evidentiary signal, then it must be the case that the other jurors have observed *G*'s and *I*'s in equal amounts. Austen-Smith and Banks assume that jurors use this information *as well* when casting a vote. They prove that if a juror's prior beliefs about the true state of the world are sufficiently strong (i.e., if *s* is sufficiently close to zero or one), and the juror uses Bayes' Rule and hypothesizes what signals other jurors must have seen if she is pivotal, then the juror maximizes her expected utility by ignoring her own evidentiary signal. In other words, her best response to everyone else voting in accordance with their evidentiary signals is not to do so. In equilibrium, the juror's vote is carried not by her observation of the trial evidence, but by the weight of her beliefs and conjectures about what others must be thinking and doing if her vote is indeed the tie-breaker.

Feddersen and Pesendorfer (1998) extend this logic to the case of unanimous verdicts. A common rationale for unanimity in juries is that it minimizes the probability of convicting the innocent. If jurors vote in accordance with their evidentiary signals, a kind of voting that Feddersen and Pesendorfer call "informative voting," then unanimity minimizes the probability of false convictions. But Feddersen and Pesendorfer identify a

Nash equilibrium in which unanimity produces more false convictions than do other decision rules because jurors need not vote informatively. In their model, if a juror is pivotal under unanimity rule, then she can infer that every other juror is voting to convict. If this is true, then she can also make an inference about how many other jurors received guilty signals. The authors identify conditions in which the weight of each juror's conjecture about what other jurors are doing leads *all of them* to conclude that they should vote to convict -- even if they all received innocent signals. False convictions come from such calculations and are further fueled by jury size (as *n* increases, so does the informational power of the conjecture "If I am pivotal, then it must be the case that every other juror is voting to convict.") Such results call into question claims about unanimity's beneficial normative properties.[7]

Driving the difference between the CJT result and newer theoretical results is the assumption that all jurors rigorously contemplate what others are doing. Questions about whether citizens think in such ways prompted clever experiments by Guernaschelli, McKelvey, and Palfrey (2000; henceforth GMP). Using students as subjects, they examined juries of different sizes (*n=3* and *n=6*). The GMP experiments lend mixed support to the recent claims. Some jurors do vote to convict despite receiving innocent signals and this behavior can lead to false convictions. But neither outcome happens as often the Nash equilibrium upon which Feddersen-Pesendorfer focus suggests. GMP (p. 416) report that where: "Feddersen and Pesendorfer (1998) imply that large unanimous juries will convict innocent defendants with fairly high probability… this did not happen

[7] Later work by Coughlan (2000) and Austen-Smith and Feddersen (2006) examines whether allowing jurors to participate in a straw poll prior to the final vote reduces the pathological effects of information aggregation identified in the focal equilibrium of Feddersen and Pesendorfer. Coughlan identifies an equilibrium where it does, but Austen-Smith and Feddersen find that this result is not robust to the introduction of inter-juror uncertainty about whether other jurors are biased for or against conviction.

in our experiment." In fact, and contrary to another conclusion from the 1998 paper, the gap between the theoretical prediction and the experimental data grew with jury size.

Before presenting our own model of such phenomena, we first review empirical research that motivates our theoretical framework. There exists a substantial psychological literature on jury decision making. It is largely grounded in experiments built around mock juries with many participants sampled from courthouse jury pools. The literature documents important attributes of how people think in jury settings. Focal citations include a series of papers and books by Nancy Pennington and Reid Hastie. Their research begins with the premise that jurors encounter a massive database of evidence during a trial. The evidence is often presented in a scrambled order. Instead of being strictly chronological, plaintiffs and defendants produce different kinds of evidence at different times. From many jurors' perspectives, the evidence is piecemeal and leaves many gaps in their attempts to understand what really happened.

How do jurors react? Pennington and Hastie explain their reactions with "story" models. Each juror attempts to make sense of the evidence by assembling it into a narrative format. A narrative comes from three sources: case-specific information acquired during the trial, a juror's knowledge of similar events, and a juror's expectations of what constitutes a complete story. Comparing the "story model" approach to other empirically-based explanations of jury decision making, MacCoun (1989: 1047) finds that it is "the only model in which serious consideration is given to the role of memory processes during the trial," while Devine, et. al. (2001:624) concludes that it is "the most widely adopted approach to juror decision making."

These studies reveal interesting variations in story content. Some jurors use complex narratives to make sense of what they see. Others use simple narratives. For our purpose, just as important is the fact that many jurors are shocked to learn of such variations after the fact. For example, Pennington and Hastie (1990: 94, emphasis added) not only found that "many jurors tended to construct *only one* of the possible stories," but also that "*jurors were surprised to discover that there were other possible stories*" that fit the evidence. Many jurors construct a simple story as a means of understanding the evidence and provide no evidence of having put any thought at all into the possibility that others drew different conclusions from the same evidence.

That jurors differ in these ways is consistent with other core findings in the psychological study of how people think. Building from studies by Cohen, et. al. (1955), Cacioppo and Petty (1982) began to document differences in how much people enjoy thinking about – and actually think about -- complex matters. While some citizens enjoy dealing with logical abstractions, others strive to minimize the mental effort devoted to such activities. Over the span of several decades, substantial variation in citizens' "need for cognition" (henceforth, NFC) have been observed (Wegener, et. al. 2000). Such variation explains and reinforces the variations in story quality observed by psychological jury scholars. Story model and need for cognition studies provide insight into the range of conjectures and beliefs on which jurors base their voting decisions.

The Next Step

At present, there is little interaction between the psychological and theoretical literatures just described. A recent quote (Hastie and Kameda 2005: 12), suggests both a reason for the isolation and a strategy for more effective interaction.

> "[GMP's] empirical study is an antidote to a previous controversial paper that argued, on the basis of a theoretical model (not behavioral data), that unanimity rule without discussion was universally inferior to the majority rule (Feddersen and Pesendorfer 1998)."

In the quote, the theory's logic is unchallenged. But the theory's relevance is called into question because it is not based on behavioral data.

To be sure, recent theoretical claims presume that jurors efficiently contemplate abstractions such as "what others must be thinking if I am pivotal." It may be the case that all jurors think in such ways or proceed "as if" they have such thoughts. But what if some do not? Are Feddersen and Pesendorfer's findings about the frequency of false convictions under unanimity rule robust to the introduction of cognitive premises that are more consistent with those seen in the story model and in the need for cognition studies?

Our model addresses this question. Like previous theory, the model's jurors are goal-oriented, in that they prefer to acquit the innocent, and strategic, in that they plan their actions to maximize their expected utility. Like previous psychological work, the model's jurors vary in how they think (or do not think) about the information that is presented to them -- and to others. To leverage the kind of variation in cognitive practices seen in the psychological work, we use the SCE concept to derive conclusions about behaviors and outcomes. As described in previous sections, we use the SCE rather than the NE because it provides a means for deriving steady state conclusions from a broad range of assumptions about how jurors think.

Our model's foundation is Feddersen-Pesendorfer (henceforth FP). It is a game with $N=\{1,2,\ldots n\}$ jurors that begins with Nature determining the state of the world. Let $\Omega=\{G,I\}$, where $\Omega=G$ means that the defendant committed the crime in question and $\Omega=I$ means that he did not. G and I occur with equal probability. No juror observes the true state of the world directly. Instead, each juror receives an *evidentiary signal*. As in previous models, each evidentiary signal is an independent Bernoulli random variable, $m_j \in \{g,i\}$, which, for each juror *j*, reveals the true value of Ω with probability $p \in (.5,1)$, and the false value of Ω with probability $1\text{-}p$. After observing m_j, each juror casts a vote $X_j \in \{A,C\}$, where $X_j=A$ is a vote by juror *j* to acquit and $X_t=C$ is a vote to convict. We focus on unanimity, so if all *n* jurors choose *C*, then the group decision is *C*; otherwise it is *A*. All jurors prefer to convict only the guilty and to set only the innocent free: $u(C,G)=u(A,I)=0$ and $u(C,I)=-q \text{ and } u(A,G)=-(1-q)$, where $q \in (0,1)$ is the same for all jurors and “characterizes a juror’s threshold of reasonable doubt” (FP 1998: 24). Juror *j*’s voting behavior is described by the strategy $\sigma_j\colon\{g,i\} \Rightarrow [0,1]$, which maps evidentiary signals into a probability of voting to convict.

We break from FP assuming that the jury contains two kinds of jurors. Some jurors are high in “need for cognition” and others are low NFC. The difference between the jurors is their ability to construct complex stories about what they do not observe and their ability to imagine that other jurors think differently.

A low NFC juror’s private signal contains her evidentiary signal along with the knowledge that unanimity is the decision rule and all jurors have identical utility functions. The private signal does not include the fact that their evidentiary signal was the result of a single draw from the Bernoulli distribution. Instead, they consider their evidentiary signal to be “the truth.” Technically, we assume that they believe that $p=1$ for all jurors. Low NFC jurors do not consider the possibility that other jurors may have received different signals. They do not think about what they cannot see. So, our low NFC jurors are like the jurors in the Pennington and Hastie studies who were shocked to learn that other jurors constructed causal stories different than their own. They also resemble the subset of actors in Hafer and Landa's (2005:9) deliberation model who are strategic because they craft strategies to attain goals but do not process information via Bayesian updating because they "do not know what they do not know."[8]

High NFC jurors differ from low NFC jurors in that their private signals are more informative. A high NFC juror's private signal includes their evidentiary signal and everything that was common knowledge in FP. Unlike their low NFC counterparts, they also know the proportion of high NFC and low NFC jurors in the jury. Therefore, they are capable of the kind of information processing assumed in the recent generation of formal models ("My vote matters only when I am pivotal and if I am pivotal, it must be the case that…"). Table 1 describes the differences between the two kinds of jurors.

	Low NFC	High NFC
Private signal permits "If I am pivotal…" thinking	No	Yes
Beliefs about *p*	*p=1* for everyone.	They know the value of *p*.
Beliefs about jury composition	Everyone is like me.	They know the number of High and Low NFC jurors.
Conjecture about others' strategies	All vote informatively.	Depends on *p*, *n*, *q* and number of high NFC jurors.
Strategic	Yes	Yes
Goal-Oriented	Yes	Yes

Table 1. Comparison between High NFC and Low NFC jurors

With this framework in hand, we use the model to reexamine the focal question of FP: With what frequency do false convictions occur? We conclude that the problem of

[8] This representation is consistent with that of Tingley (2005). In reviewing work by Byrne et al. (2000), he highlights "actions (as opposed to inactions)" as being likely sources for counterfactual generation.

false convictions increases with the proportion of high NFC jurors. When all jurors are low NFC, even "strategic" voting under unanimity rule minimizes the frequency of false convictions.

To reach this conclusion, we make two additional assumptions. First, like FP, we focus on "symmetric" and "responsive" equilibria.[9] *Symmetry* requires that similarly situated actors take identical actions. So, we characterize equilibria where all low NFC jurors choose identical strategies and all high NFC jurors choose identical strategies. Since high NFC and low NFC jurors are not similarly situated – they receive different private signals – our symmetry requirement does not require the two groups to choose identical strategies. *Responsiveness* requires that jurors change their vote as a function of their evidentiary signal with positive probability (i.e., $p\sigma_j(g) +(1-p)\ \sigma_j(i) \neq (1-p)\sigma_j(g)+p\sigma_j(i))$. Second, we follow the common practice of eliminating weakly dominated strategies from consideration.

We begin with the case where all jurors are low NFC. To determine whether a particular set of strategies constitutes a SCE, we must determine whether a juror's observations are consistent with her conjecture and beliefs.

Low NFC Proposition. *If all jurors are low NFC, then all jurors voting informatively is the only responsive and symmetric SCE.*

> **Proof:** Every juror believes that all other jurors see the same signal. If a juror j observes $m_j=G$, then she believes that $\Omega=G$ with probability 1. Given the knowledge that all jurors have identical utility functions, she conjectures that all other jurors are voting to convict. If $\sigma_j(g)=1$ (she votes to convict), then her belief and conjecture lead her to expect utility $u(C,G)=0$. If $\sigma_j(g)=1-z,\ z\in[0,1]$, then she conjectures that her vote will preclude a unanimous guilty verdict with probability z. Given her belief and conjecture, she expects utility $u(A,G)=-z(1-q)$. Since $q\in(0,1)$, her expected utility is maximized at $z=0$. Therefore, if $m_j=G$, then any

[9] Other equilibria exist, including all voters choosing to acquit regardless of their signal. This is true for both FP's NE-based inferences and our SCE-based inferences.

> responsive, symmetric SCE must include $\sigma_j(g)=1, \forall j$. If $m_j=I$, then the juror believes that $\Omega=I$ and conjectures that all other jurors are voting to acquit. Whether she votes to convict or acquit, the defendant will be acquitted. Given her belief and conjecture, she expects utility $u(A,I)=0$ from any strategy $\sigma_j(i) \in [0,1]$, however, only $\sigma_j(i)=0$ survives weak domination. **Q.E.D.**

In this case, informative voting constitutes equilibrium behavior. This outcome is unlike the Nash-based conclusion where all jurors are presumed to have common beliefs and conjectures. Moreover, in the SCE case, the false conviction probability is $(1-p)^n$, as is true in the CJT. In other words, a false conviction occurs only if every juror receives a false "guilty" signal when the true state of the world is innocent. If we take the normal size of the jury (n=12) and use the least-flattering assumption about signal quality in the theoretical papers cited (p approaches .5 from above), then the probability of a false conviction is roughly 1/4096. As signal quality or jury size increase, the probability of a false conviction goes to zero. We now consider the case where all jurors are high NFC.

High NFC Proposition: *Under the technical conditions described in Feddersen and Pesendorfer's Proposition 2(1998: 26), if all jurors are high NFC then the only responsive and symmetric SCE entails $\sigma(g)=1$ and $\sigma(i)>0$.*

Here, the unique symmetric and responsive SCE is identical to FP's unique and responsive NE. The proof follows accordingly and in the case described above ($p \approx .5$) the probability of a false conviction diverges away from zero as jury size increases.

Therefore, it is not strategic voting *per se* that generates FP's high rate of false convictions – as goal-oriented low NFC jurors in our model who are attempting to maximize their expected utility do not generate high false conviction rates. Driving the increase in false convictions is the assumption that jurors process information in a particular way. False convictions are caused by the presence of jurors who conjecture that all other jurors are thinking in the same, complex manner as they are.

To consider what these results imply for the normative qualities of unanimous verdicts with real juries, we recall from the psychological literature that most juries will likely contain a mix of high and low NFC jurors. In our model, the two kinds of jurors can be mixed in many different proportions, but a full mathematical treatment of behavior in all such cases is beyond the scope of this example. We can, however, use the results derived above to give some intuition about how the presence of jurors who vary in the kinds of stories they construct affects the probability of false convictions.

Suppose that there exists a jury containing 1 high NFC juror and $n-1$ low NFC jurors and, as in FP's focal example, let $p=.7$ and $q=.5$. For low NFC jurors, this case is observationally equivalent to that described in the "Low NFC Proposition." Therefore, any symmetric and responsive SCE must involve all such jurors voting informatively. Moreover, if $n>2$, then this SCE includes the high NFC juror voting to convict regardless of their evidentiary signal.[10]

What is the probability of false convictions in this case? As Figure 2 shows, it is far less than reported in FP. In our version, the probability of a false conviction is $(1-p)^n p$. This probability is lower than FP's because only a limited number of jurors vote contrary to their evidentiary signals. In FP, symmetry requires that if one juror votes against his evidentiary signal with positive probability, then all other jurors must do the same. This attribute of FP's example is what drives the false conviction probability away from zero as jury size grows. In our version of the example, letting high and low NFC jurors think in different ways drives this same probability to zero very quickly as they jury grows.

[10] If the high NFC juror receives an innocent signal, he calculates the probability of guilt as $Z=((1-p)p^{n-1})/((1-p)p^{n-1} + p(1-p)^{n-1}))$ and votes to convict if $-q(1-Z) > -(1-q)Z$. When $p=0.7$ and $q=0.5$, this inequality is satisfied for $n > 2$. If he receives a guilty signal, he calculates the probability of guilt as $Z'=p^n/(p^n + (1-p)^n)$ and votes to convict if $-q(1-Z') > -(1-q)Z'$. For $p=0.7$ and $q=0.5$, this inequality is satisfied for all n.

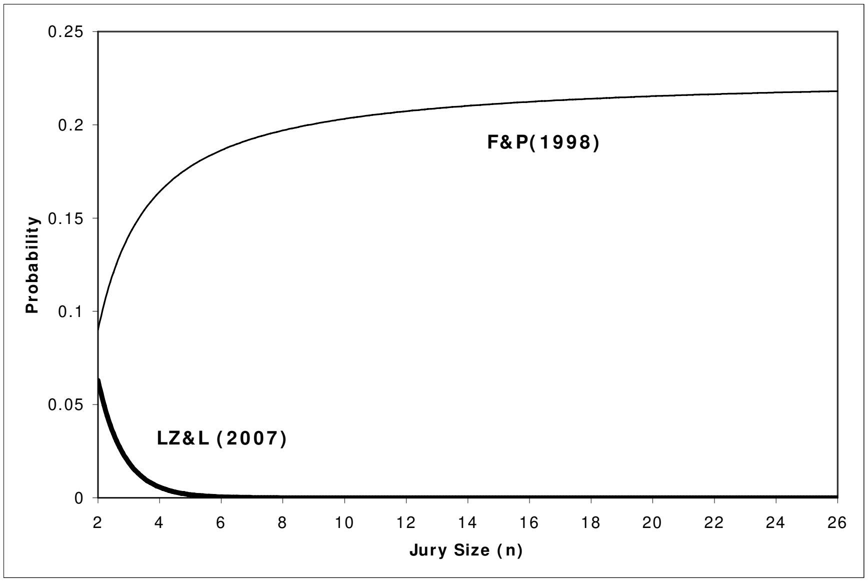

Figure 2. The Probability That an Innocent Defendant is Convicted as a Function of Jury Size for *p=.7* and *q=.5*.

More generally, the extent to which the pathologies of unanimity rule pointed out by FP occur in our model is a function of the ratio of high NFC to low NFC jurors. When all jurors are low NFC, unanimity rule retains the beneficial normative properties attributed to it by the CJT. As high NFC jurors appear, so does the probability of false convictions.

Such results imply that to understand how often unanimity rule convicts the innocent, we need to know more about how jurors think. In particular, we should examine questions such as "Under what conditions are *w* of *N* jurors likely to act like high NFCs?" For cases where most jurors are like low NFCs, our model suggests that unanimity rule will not generate many false convictions. But where evidence suggests that most or all jurors are high NFCs who contemplate information in ways described in the recent generation of game theoretic models, we would follow Feddersen and Pesendorfer in questioning the virtues of unanimous verdicts.

Discussion: SCE and QRE

Viewing jury decision making through the conceptual lens that the SCE concept promotes is complementary to the Quantal Response Equilibrium (QRE) approach adopted by Guernaschelli, McKelvey, and Palfrey. Both concepts address the empirical challenges caused by the gap between actual human reasoning and that posited in Nash-based concepts.

SCE challenges scholars to think about the contours of reasoning before the data analysis is conducted. In this example, we relied on the psychological jury literature to inform assumptions about juror beliefs and conjectures. This linkage led us to focus on a set of theoretical conclusions that are easier to reconcile with observed behavior in psychology-based jury studies.

QRE challenges scholars to deal with the problem in a different way. Nash-based behavior is assumed, and statistical procedures are used to estimate the shape of an error function. So, in the GMP paper, the QRE does not provide an *ex ante* prediction about behavior that is superior to FP's NE prediction, but it does provide the basis for a statistical analysis of the data from which a stochastic error term is derived. Once the error term is fed back into the theoretical analysis, GMP's improved explanation emerges.

The SCE and QRE concepts challenge researchers to deal with the psychological underpinnings of strategic behavior explicitly and transparently. Whether SCE, QRE, or a Nash-based equilibrium concept is most appropriate for studying juries is an interesting question. We contend that such questions are, at least in part, empirical. Some people are probably better characterized by a SCE-based algorithm and others by a QRE-based approach. There is no reason that empirical inquiries into which concept's representation of human reasoning best fits the reasoning styles of the actors to be modeled should not inform the choice of solution concept.

For the case of jury decision making, it is worth noting that both QRE and SCE can explain GMP's observation of a widening gap between the theoretical predictions and the experimental observations as jury size grows. GMP treat the gap as a result of respondents making errors in their attempts to implement NE strategies. Our SCE-based explanation is that as jury size grows, the cognitive effort required to act like a high NFC voter (If I am pivotal, …) grows. Faced with growing complexity, and holding motivation constant, jurors are more likely to seek simple stories of cause-and-effect – they are more likely to act like low NFC jurors. Therefore, the gap between the probability of false convictions and the observed rate of false convictions should grow with jury size.

Conclusion

Common game-theoretic equilibrium concepts entail implicit assumptions about how people reason. One assumption is that people think about others in potentially complex ways. Another is that their reasoning is similar in important respects. But evidence from psychology and political science make it unlikely that all political actors think in such ways.

This paper responds to those findings. As leading empirical scholars make substantive contributions through methodological work that reconciles properties of statistical estimators with properties of focal data (see, e.g., Aldrich and McKelvey 1977, King 1989, King, Keohane, and Verba 1994), analogous attempts to reconcile equilibrium concepts with properties of focal data can constitute methodological advances that allow scholars to draw more effective and credible inferences.

To be sure, implementing SCE poses new challenges. On the one hand, it allows us to be much more explicit about the kinds of conjectures upon which actors base their strategies. On the other hand, if we want to reduce the number of focal equilibria, then the SCE approach induces us to provide a more detailed psychological account than is true for many Nash-based approaches. For some, the psychological black boxes whose theoretical implications an SCE can bring to light will appear to be a stack of Pandora's boxes – ones that are just as well left closed. We disagree. The SCE concept does not create the problem of whether any particular model represents cognition accurately, it only makes consequences of ignoring cognition apparent (see, e.g., de Figueiredo, Weingast, and Rakove 2006 and Tingley 2005).

In response to these challenges, scholars can benefit from asking informed, direct, and concrete questions about how the actors they model view their environs and those around them. Psychology is producing a growing range of findings about the kinds of information to which political actors attend and remember (see, e.g., Kuklinski 2001, 2002). Such information can play an important role in clarifying the conditions under which key political actors run cognitive assessments of any particular quality.

Many political actors pay very limited attention to, or have very different beliefs about, political phenomena. They may not observe others' actions or know much about others' payoffs. But they can still attempt to better their situations by strategically conditioning their actions on what they observe. We offer this paper as a constructive starting point for developing formal models that are easier to reconcile with emergent psychological phenomena.

References

Aldrich, John F., and Richard D. McKelvey. 1977. “A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections.” *American Political Science Review* 71(1): 111-130.

Aumann, Robert. 1974. “Subjectivity and Correlation in Randomized Strategies.” *Journal of Mathematical Economics* 1(1): 67-96.

Aumann, Robert and Adam Brandenberger. 1995. “Epistemic Conditions for Nash Equilibrium.” *Econometrica* 63(5): 1161-1180.

Austen-Smith, David, and Jeffrey S. Banks. 1996. “Information Aggregation, Rationality, and the Condorcet Jury Theorem.” *American Political Science Review* 90(1): 34-45.

Austen-Smith, David, and Timothy J. Feddersen. 2006. “Deliberation, Preference Uncertainty, and Voting Rules.” *American Political Science Review* 100(2): 209-217.

Bernheim, B. Douglas. 1984. “Rationalizable Strategic Behavior.” *Econometrica* 52(4): 1007-1028.

Bjork, Elizabeth Ligon, and Robert Bjork. 1996. “Continuing Influences of To-Be-Forgotten Information.” *Consciousness and Cognition* 5(1-2): 176-196.

Cacioppo, John T., and Richard E. Petty. 1982. “The Need for Cognition.” J*ournal of Personality and Social Psychology* 42(1): 116-131.

Camerer, Colin F., George Loewenstein, and Matthew Rabin. 2003. *Advances in Behavioral Economics*. New York: Russell Sage Foundation.

de Figueiredo, Rui, Jack Rakove, and Barry R. Weingast. 2006. “Rationality, Inaccurate Mental Models and Self-Confirming Equilibrium: A New Understanding of the American Revolution.” *Journal of Theoretical Politics* 18(4): 384-415.

Dekel, Eddie, Drew Fudenberg, and David K. Levine. 1999. “Payoff Information and Self-Confirming Equilibrium.” *Journal of Economic Theory* 89(2): 165-185.

Dekel, Eddie, Drew Fudenberg, and David K. Levine. 2004. “Learning to Play Bayesian Games.” *Games and Economic Behavior* 46(2): 282-303.

Devine, Dennis J., Laura D. Clayton, Benjamin B. Dunford, Rasmy Seying, and Jennifer Pryce. 2001. “Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups.” *Psychology, Public Policy, and Law* 7(3) 622-727.

Feddersen, Timothy, and Wolfgang Pesendorfer. 2006. “Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting.” *American Political Science Review* 92(1): 23-35.

Fudenberg, Drew, and David M. Kreps. 1995. “Learning in Extensive-Form Games I: Self-Confirming Equilibrium.” *Games and Economic Behavior* 8(1): 20-55.

Fudenberg, Drew, and David K. Levine. 1993. “Self-Confirming Equilibrium.” *Econometrica* 61(3): 523-545.

Fudenberg, Drew, and David K. Levine. 1998. *The Theory of Learning in Games*. Cambridge, MA: MIT Press.

Hafer, Catherine, and Dimitri Landa. 2006. “Deliberation as Self-Discovery and Institutions for Political Speech.” Forthcoming, *Journal of Theoretical Politics* 19.

Harsanyi, John. 1967. “Games with Incomplete Information Played by `Bayesian’ Players I: The Basic Model.” *Management Science* 14(3): 159-182.

Harsanyi, John. 1968. “Games with Incomplete Information Played by `Bayesian’ Players II: Bayesian Equilibrium Points.” *Management Science* 14(3): 320-334.

Hastie, Reid, and Tatsuya Kameda. 2005. “The Robust Beauty of Majority Rules in Group Decisions.” *Psychological Review* 112(2)”494-508.

Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell. 1995. *Essentials of Neural Science and Behavior*. Norwalk, CT: Appleton and Lange.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Kreps, David, and Robert Wilson. 1982. “Sequential Equilibria.” *Econometrica* 50(4): 863-894.

Kuklinski, James H. (Ed.) 2001. *Citizens and Politics: Perspectives from Political Psychology*. New York: Cambridge University Press.

Kuklinski, James H. (Ed.) 2002. *Thinking About Political Psychology*. New York: Cambridge University Press.

Lupia, Arthur, Mathew D. McCubbins, and Samuel L. Popkin. 2000. “Beyond Rationality: Reason and the Study of Politics.” In Arthur Lupia, Mathew D. McCubbins, and Samuel L. Popkin (eds.) *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*. New York: Cambridge University Press.

MacCoun, Robert J. 1989. “Experimental Research on Jury Decision Making.” *Science* 244(4908): 1046-1049.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. “Quantal Response Equilibria in Normal Form Games.” *Games and Economic Behavior* 10(1): 6-38.

Nash, John. 1950. "Equilibrium Points in n-Person Games." *Proceedings of the National Academy of Sciences* 36(1): 48-49.

Pearce, David G. 1984. "Rationalizable Strategic Behavior and the Problem of Perfection." *Econometrica* 52(4): 1029-1050.

Pennington, Nancy, and Reid Hastie. 1990. "Practical Implications of Psychological Research on Juror and Jury Decision Making." *Personality and Social Psychology Bulletin* 16(1): 90-105.

Pennington, Nancy, and Reid Hastie. 1993. "Reasoning in Explanation-based Decision Making." *Cognition* 49(1-2): 123-163.

Rubinstein, Ariel. 1998. *Modeling Bounded Rationality.* Cambridge, MA: MIT Press.

Satz, Debra, and John Ferejohn. 1994. "Rational Choice and Social Theory." Jo*urnal of Philosophy* 91(2): 71-87.

Schacter, Daniel L. 1996. *Searching for Memory: The Brain, The Mind, and The Past.* New York: Basic Books.

Schacter, Daniel L. 2001. *The Seven Sins of Memory: How the Mind Forgets and Remembers.* Boston: Houghton-Mifflin.

Tingley, Dustin. 2005. "Self-confirming Equilibria in Political Science: Cognitive Foundations and Conceptual Issues." Paper presented at the 2005 Annual Meeting of the American Political Science Association, Philadelphia, PA.

Turner, Mark. 2000. "Backstage Cognition in Reason and Choice." In Arthur Lupia, Mathew D. McCubbins, and Samuel L. Popkin (eds.), *Elements of Reason: Cognition, Choice and the Bounds of Rationality*. New York: Cambridge University Press.

Turner, Mark. 2001. *Cognitive Dimensions of Social Science*. Oxford: Oxford University Press.

Wegener, Duane T., Norbert L. Kerr, Monique A. Fleming, and Richard E. Petty. 2000. "Flexible Corrections of Juror Judgments: Implications for Jury Instructions." *Psychology, Public Policy, and Law* 6(3):629-654.