A Cognitive Model of Strategic Deliberation and Decision Making

Russell Golman (rgolman@andrew.cmu.edu)
Carnegie Mellon University, Pittsburgh, PA.

Sudeep Bhatia (bhatiasu@sas.upenn.edu)
University of Pennsylvania, Philadelphia, PA.

Abstract

We study game theoretic decision making using a bidirectional evidence accumulation model. Our model represents both preferences for the strategies available to the decision maker, as well as beliefs regarding the opponent's choices. Through sequential sampling and accumulation, the model is able to intelligently reason through two-player strategic games, while also generating specific violations of Nash equilibrium typically observed in these games. The main ingredients of accumulator models, stochastic sampling and dynamic accumulation, play a critical role in explaining these behavioral patterns as well as generating novel predictions.

Keywords: Decision making; Game theory; Sequential sampling; Preference accumulation

Introduction

Game theory studies the behavior of idealized decision makers. The standard solution concept for a strategic game is Nash equilibrium, which relies on common rationality and accurate expectations. Given expectations of others' choices, players behave rationally, and the resulting play conforms to these expectations (Luce & Raiffa, 1957).

Not surprisingly, human decision makers display numerous systematic departures from Nash equilibrium (see Camerer, 2003 for a review). We present a cognitive model of strategic deliberation and choice in one-shot, two-player games, that is able to accommodate these departures. Our model proposes that decision makers dynamically and stochastically accumulate both their own preferences for available strategies, as well as beliefs about the opponent's preferred strategies. There are bidirectional relationships between preferences and beliefs, so that beliefs about what the opponent will choose influence the decision makers' preferences, and these preferences in turn influence beliefs about the opponent's choices. Ultimately, decision makers can respond to what they think the opponent will do, and also revise these beliefs as they deliberate.

Our model can be seen as an extension of decision field theory (Busemeyer & Townsend, 1993; also Bhatia, 2014 and Rieskamp, 2006), an existing accumulator-based theory of non-strategic risky choice. Accumulator models rely on two main ingredients: stochastic sampling and dynamic accumulation (see Busemeyer, 2015 for a review). These ingredients are critical in our model for making deliberation subject to intrinsic variability and requiring it to play out over time, and we show that both ingredients have a central role in capturing the behavioral patterns observed in strategic choice. By demonstrating the relationship between our model and established preference accumulation models, we demonstrate that a single framework can be used to understand choice behavior across a variety of non-strategic and strategic settings.

Game Theoretic Decision Making

In strategic games, two or more players make choices over a set of strategies. Crucially, the strategies chosen by the players collectively determine the outcomes of the game, so that each player's utility depends on the other's choice as well as on their own. We define a finite-strategy two-player game with a set of pure strategies for each player, $S_1 = \{s_{11}, \dots s_{1N}\}$ and $S_2 = \{s_{21}, \dots s_{2M}\}$ respectively, and a pair of payoff functions u_1 and u_2 that give each player's utility for each profile of pure strategies (s_{1i}, s_{2j}). Thus if player 1 selects s_{1i} and player 2 selects $s_{2j}$the utility for player 1 is $u_1(s_{1i}; s_{2j})$ and the utility for player 2 is $u_2(s_{2j}; s_{1i})$, with $\boldsymbol{u}_{ij} = \left(u_1(s_{1i}; s_{2j}), u_2(s_{2j}; s_{1i})\right)$. We define the set of best responses for player μ to an opponent's strategy $s_{-\mu}$ as $\mathrm{BR}\left(s_{-\mu}\right) = \arg\max u_{\mu}(s_{\mu};\, s_{-\mu})$. Then a pure strategy Nash equilibrium can be defined as a strategy profile (s_{1i}, s_{2j}) such that $s_{1i} \in \mathrm{BR}(s_{2j})$ and $s_{2j} \in \mathrm{BR}(s_{1i})$.

There are a number of settings where Nash equilibrium fails to accurately describe human behavior. For example, Nash equilibrium predicts unraveling when players have incentives to undercut each other. Consider the traveler's dilemma game (Basu, 1994), in which two travelers have lost identical items and must request compensation. The airline accepts the lower claim as valid and pays that amount to both players, and, additionally penalizes the higher claimant with a fee and rewards the lower claimant with a bonus. We represent this game with the strategy sets $S_1 = S_2 = \{20,30,\dots,90\}$, where x_{1i} and x_{2j} correspond to the amounts (in dollars) associated with strategies s_{1i} and s_{2j}, and we have utilities $\boldsymbol{u}_{ij} = (0.01(x_{2j} - \gamma), 0.01(x_{2j} + \gamma))$ if $x_{1i} > x_{2j}$, $\boldsymbol{u}_{ij} = (0.01x_{1j}, 0.01x_{2j})$ if $x_{1i} = x_{2j}$, and $\boldsymbol{u}_{ij} = (0.01(x_{1j} + \gamma), 0.01(x_{2j} - \gamma))$ if $x_{1i} < x_{2j}$. Here γ corresponds to the reward/penalty offered by the airline, and is set so that $10 < \gamma \leq 20$. For comparability with other games, we have scaled utilities to lie between 0 and 1.

The airline's scheme rewards undercutting the other traveler. The best response is always to claim exactly 10 less than the other traveler does (if it is feasible to do so).

As a result, the only Nash equilibrium strategy for both players is to claim 20. In experiments average claims actually are well above the lower bound that Nash equilibrium predicts (e.g. Capra et al., 1999).

Experiments on the traveler's dilemma game also find that claims are higher when the reward/penalty, γ, is lower. This payoff sensitivity is hard to reconcile with players choosing best responses to the strategies they expect their opponent to play. Nash equilibrium predicts that responses in the traveler's dilemma should be independent of γ, as changing payoffs without changing best responses should have no effect on choice behavior.

Another setting in which Nash equilibrium fails to appropriately describe behavior involves coordination games. These are games with multiple pure strategy Nash equilibria, in which players are incentivized to choose the same strategy. Due to the presence of multiple equilibria, Nash theory cannot make precise predictions. However, human decision makers are often fairly predictable. Consider the Hi-Lo coordination game, in which decision makers have to choose between two strategies: Hi and Lo. In this game we have: $\boldsymbol{u_{ij}}$ = (1.0,1.0) if both players both choose Hi; $\boldsymbol{u_{ij}}$ = (γ, γ), with $0 < \gamma < 1.0$, if both plays choose Lo; and $\boldsymbol{u_{ij}}$ = (0,0) if they choose different strategies. Not surprisingly, decision makers almost always successfully coordinate on Hi-Hi to obtain the highest possible rewards in this game (Colman, 2003).

In some games, decision makers do not choose any of the Nash equilibrium strategies when the potential costs of miscoordination are too great. This can be observed in the boobytrap game, which is a standard prisoner's dilemma augmented with a third option that allows decision makers to purchase a "boobytrap" to punish their opponent if he or she defects (Misyak & Chater, 2014). Particularly, we have $\boldsymbol{u_{ij}}$ = (0.9,0.9) if both players cooperate, $\boldsymbol{u_{ij}}$ = (0.8,0.8) if both players defect, and $\boldsymbol{u_{ij}}$ = (0.89,0.89) if both players choose boobytrap. Additionally, $\boldsymbol{u_{ij}}$ = (0.7,1) if player 1 cooperates and player 2 defects, $\boldsymbol{u_{ij}}$ = (0.9,0.89) if player 1 cooperates and player 2 chooses boobytrap, and $\boldsymbol{u_{ij}}$ = (0,0.69) if player 1 defects and player 2 chooses boobytrap (and vice versa, as the game is symmetric). Nash equilibrium predicts that decision makers should ignore the boobytrap choice, however the presence of the boobytrap greatly increases the rate of cooperation in the game, contradicting the prediction of Nash equilibrium.

Yet another set of findings not accounted for by Nash equilibrium theory involves strategy salience. In many games, strategies with salient labels are more likely to be chosen. This is the case in coordination games offering multiple payoff identical strategies, with one of the strategies circled, underlined, or made salient using some other technique. Here players can coordinate successfully by selecting the salient strategy (Mehta et al., 2004).

Bidirectional Accumulation

We propose an extension to a preexisting accumulator model of risky choice, decision field theory (Busemeyer & Townsend, 1993). As in decision field theory, decision makers use two layers of nodes: one to accumulate preferences in favor of the available choice options, and one to represent the probabilistic events involved in the decision. In the strategic context, the choice options are the strategies available to the decision maker and the events are the possible strategies the opponent may use. Thus, the strength of the connection from the node representing a strategy *j* for the opponent to the node representing preference for a decision maker's strategy *i*, is proportional to the utility of strategy *i* for the decision maker, given that the opponent plays strategy *j*. Decision makers sample the events according to the subjective probabilities they assign to their occurrence. Thus, strategies that are more likely to be played by the opponent are sampled more frequently and thereby play a larger role in determining the decision makers' preferences.

Decision field theory assumes that decision makers' beliefs about events (and subsequently sampling probabilities for these events) are fixed. For the most part this is reasonable: decision makers' preferences do not influence the actual probability with which different events occur. This assumption is less reasonable in strategic settings. Sophisticated opponents, who can anticipate decision makers' choices, will adjust their own choices to maximize their reward. We thus assume a bidirectional accumulation process to represent strategic deliberation. At each time period, decision makers sample one of their opponent's strategies based on the activations of the nodes corresponding to these strategies, and update their preferences over their own strategies based on this sample. Decision makers then sample one of their own strategies based on the activation of the nodes, and use this sample to update their beliefs about their opponent's choices. In essence, decision makers have dynamically changing mental representations for not only their own preferences, but also their beliefs about their opponents' preferences, allowing them to deliberate intelligently using perspective taking and a sophisticated theory of mind.

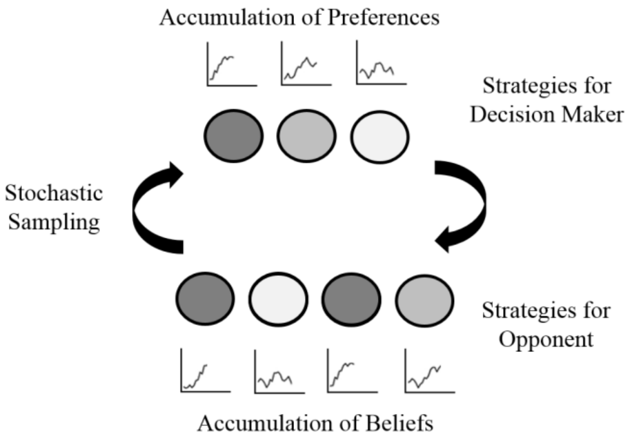

Figure 1: Illustration of bidirectional accumulation model

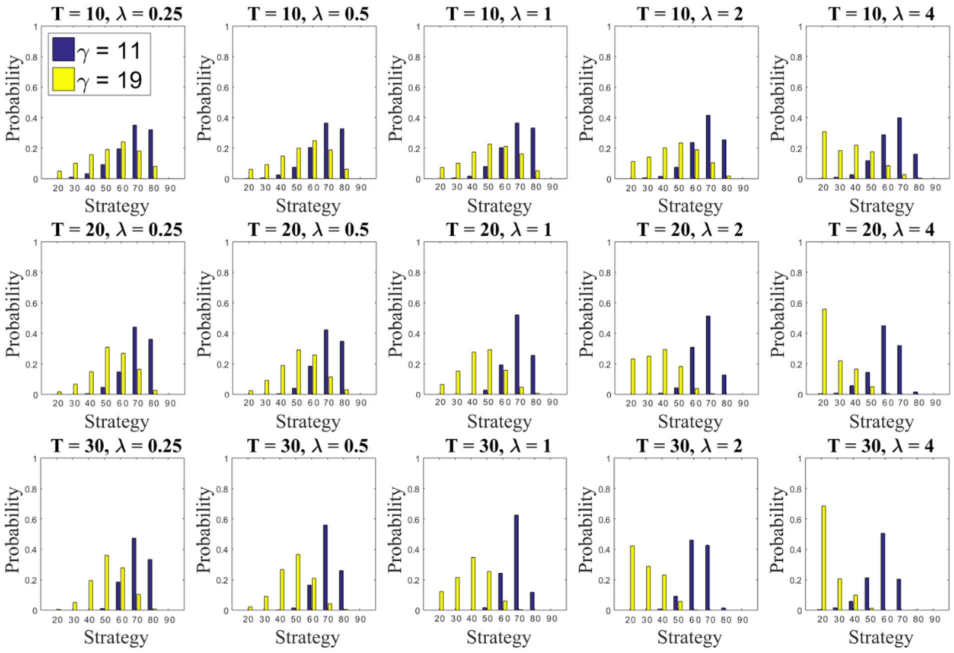

Figure 2: Simulated distribution of choices in the traveler's dilemma.

Formally, if the decision maker has to choose from the set of strategies $S_1 = \{s_{11}, \dots s_{1N}\}$, then the preference layer in our model consists of *N* nodes, with node *i* representing strategy s_{1i}. The activation of node *i* at time *t*, $A_{1i}(t)$ corresponds to the decision maker's preference for strategy *i* at time *t*. Correspondingly if the opponent has the set of available strategies $S_2 = \{s_{21}, \dots s_{2M}\}$, then the belief layer in our model consists of *M* nodes, with node *j* representing strategy s_{2j}. The activation of node *j* at time *t*, $A_{2j}(t)$ corresponds to the beliefs that the decision maker has about the opponent's preference for strategy *j*, at time *t*. We also denote the salience bias of any strategy *i* (for the decision maker) or *j* (for the opponent) as σ_{1i} or σ_{2j}. These salience biases σ_{1i} and σ_{2j} are independent of the decision process and are determined by various exogenous factors.

At each time period *t*, the decision maker draws one sample of the opponent's strategies. We assume that a softmax (logit) function, with stochasticity parameter $\lambda > 0$, determines the effect of activation strength and the exogenous salience bias on sampling probability. Thus, the probability of sampling strategy *j* at time *t* is given by: $p_j = e^{\lambda(A_{2j}(t-1)+\sigma_{2j})} / \sum_{k=1}^{M} e^{\lambda(A_{2k}(t-1)+\sigma_{2k})}$. If the opponent's strategy *j* is sampled, then the decision maker observes the utility for each strategy *i* conditional on the opponent playing this sampled strategy: $u_1(s_{1i}; s_{2j})$. The decision maker's preferences are then updated based on this calculated utility, so the activation for each strategy *i* becomes: $A_{1i}(t) = A_{1i}(t-1) + u_1(s_{1i};\ s_{2j})$.

As discussed, beliefs about the opponent's strategies are themselves updated based on the utility the opponent would derive conditional on a sample of the decision maker's strategies. Thus, after updating activation states $A_{1i}(t)$, decision makers draw one sample of their own strategies. The probability of sampling strategy *i* at time *t* is given by: $q_i = e^{\lambda(A_{1i}(t)+\sigma_{1i})} / \sum_{k=1}^{N} e^{\lambda(A_{1k}(t)+\sigma_{1k})}$. After sampling strategy *i*, the updated activation for each opponent strategy *j* is $A_{2j}(t) = A_{2j}(t-1) + u_2(s_{2j};\ s_{1i})$.

The deliberation process begins with nodes having no initial activation: $A_{1i}(0) = 0$ for all *i*; $A_{2j}(0) = 0$ for all *j*. Activation accumulates according to these equations until a time $t = T$. At this time, the most preferred strategy --that is, the one whose node has the highest activation-- is the strategy that is chosen by the decision maker. The parameter *T* corresponds to an exogenous time limit on the deliberation process, and represents the amount of time taken by the decision makers to make their choices. The proposed model is illustrated in Figure 1.

Explaining Behavioral Findings

In order to demonstrate how our model works, we use it to simulate choices in the games we introduced earlier. Our simulations use the same strategy and reward profiles as in examples in the previous section. For each of the games and each set of parameter values, we simulate our model 3000 times and report aggregate choice probabilities. We find that the model is fairly robust to parameter variation in the range $\lambda \in [0.25, 4]$ and $T \in [10, 30]$, and any combination of parameter values in this range produces behavior consistent with the empirical findings we have reviewed. When not explicitly specified, we set salience to $\sigma_{1i} = \sigma_{2j} = 0$.

Traveler's Dilemma. In the traveler's dilemma our model predicts a failure of unraveling. This is demonstrated in Figure 2 which plots the probability of selecting strategies in the set {20, 30, …, 90} for $\gamma = 11$ and $\gamma = 19$, with varying values of λ and T. Instead of predicting that players always claim the lowest possible amount, as in Nash equilibrium, here the model generates a distribution of choices that spreads across the range of strategies available to the decision maker. The model also displays payoff sensitivity. For a larger value of the reward/penalty parameter ($\gamma = 19$), the distribution of choices is smaller.

The intuition behind the model's predictions is appealing. For low rewards/penalties, i.e. low values of γ, the payoffs when both players make high claims are significantly higher than the payoffs when there is a low claim. The potential cost of missing out on this high payoff dwarfs the cost of making a higher claim than the opponent or the benefit of making a lower claim than the opponent. So, a few samples (or even a single sample) of the opponent playing a high claim will lead to high activation for one's own high claims. As beliefs about the opponent's strategy are updated, there will be more samples of high claims, and strategies involving an additional step of undercutting can accumulate the most utility. The number of steps of undercutting that does occur depends on payoff magnitudes. Increasing the parameter γ encourages undercutting. Although it does not affect best responses (that is, the ranking of payoffs in any given sample of play), it does affect the accumulation of payoffs over time, so strategies involving more undercutting can accumulate activation more quickly.

Stochastic sampling plays an important role in the emergence of payoff sensitivity. The magnitudes of payoff differences affect the probabilities of sampling each strategy. The degree of responsiveness to the payoff parameter γ that we observe in the predicted choices for this game depends on the logit sampling parameter λ. Comparing across the columns of Figure 2, we see larger shifts in the distribution of choices from a change in the reward/penalty parameter γ as the parameter λ increases.

Our model also makes new predictions about the relationship between decision time and the strategy chosen in the traveler's dilemma. Each step of undercutting takes time, and thus both the decision maker's preferred claim and the beliefs about the opponent's claim should thus decrease over time. Comparing across the rows of Figure 2, we observe lower claims when the decision time T is larger. Indeed, experiments have revealed that decision makers take longer to choose the lowest claim than the highest claim (Rubinstein, 2007).

Overall, with reasonable parameter values, the model predicts a failure of unraveling. Indeed, full unraveling, consistent with Nash equilibrium would only occur with very large values of λ and T, i.e., when poorly performing strategies are rarely sampled and there are many periods of sampling and iterative updating. Assuming deterministic sampling of best responses or unlimited decision time would thus lead to poor behavioral predictions for the traveler's dilemma. Conversely, assuming uniformly random sampling would lead to unreasonably high odds of choosing 80 relative to 70, underestimating people's ability to put themselves in their opponents' shoes and think strategically about their responses.

The Hi-Lo Game. Although the Hi-Lo game has two Nash equilibria, our model favors the Hi-Hi equilibrium. This is shown in Figure 3, which plots the probability of choosing Hi as a function of the payoff for coordinating on Lo (γ) for varying values of T and λ. Across all parameter values we consider, Hi is the modal choice. When the payoff asymmetry is extreme, i.e., $\gamma = 0.1$, Hi is almost certain to be chosen. Still, as the Lo-Lo payoff γ increases, so does the probability of choosing Lo.

Predictable coordination in the Hi-Lo game is intuitive. The Hi strategy, which offers higher payoffs in the case of successful coordination, accumulates more activation when it is sampled from the other layer of the network than the low strategy does. This creates a feedback effect, so the model is more likely to think about Hi when forming beliefs about the opponent's choices. Believing that the opponent will choose Hi further reinforces the model's preference Hi.

The positive feedback loop, along with stochastic sampling, actually facilitates the occasional choices of Lo. If Lo is sampled first, it gains an advantage, and it becomes more likely to be sampled again. As the logit sampling parameter λ increases, it becomes somewhat more likely (albeit still not very likely) that the model repeatedly samples Lo early on, gets fixated on this strategy, and eventually chooses it. In the extreme case that the sampling parameter λ gets unrealistically large, the strategy sampled in the first time period may be sampled forever thereafter, completely determining the path of the deliberation. Since both strategies have the same probability of being sampled in the first period of the deliberation, the model's choice distribution approaches a 50-50 split between Hi and Lo independent of γ for very large values of λ. As can be seen, decision time has little effect on the choice distribution, with longer deliberation only slightly reducing noise and increasing the probability of selecting the modal choice, Hi.

The Boobytrap Game. Our model deviates far from Nash equilibrium in the boobytrap game as well. For $\lambda \in [0.25, 4]$ and $T \in [10, 30]$, it predicts that players will almost certainly cooperate (cooperation with a greater than 90% chance for all parameters). Here, a non-Nash strategy is favored due to the high magnitude of its advantage when the other player does not best respond compared to the low magnitude of its cost when the other player does respond rationally. Against the boobytrap strategy, defection is extremely undesirable. The model predicts that players will never choose the boobytrap strategy, because it is dominated by cooperation. However, the model predicts that decision makers usually will contemplate this boobytrap strategy as part of their deliberation, and this causes their preferences for defection to drop strongly.

Again, our model's behavior would be very different with an assumption of deterministic sampling of the most highly activated strategy. With deterministic sampling, the model is confident that the boobytrap strategy will not be played, so it chooses to defect.

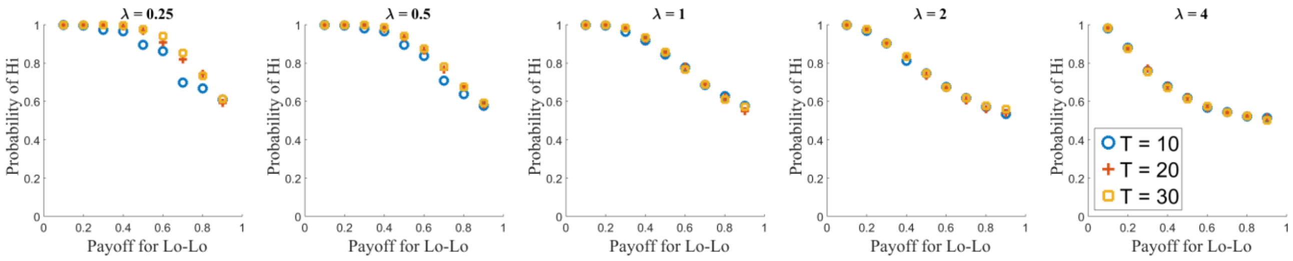

Figure 3: Simulated probability of choosing Hi in the Hi-Lo game.

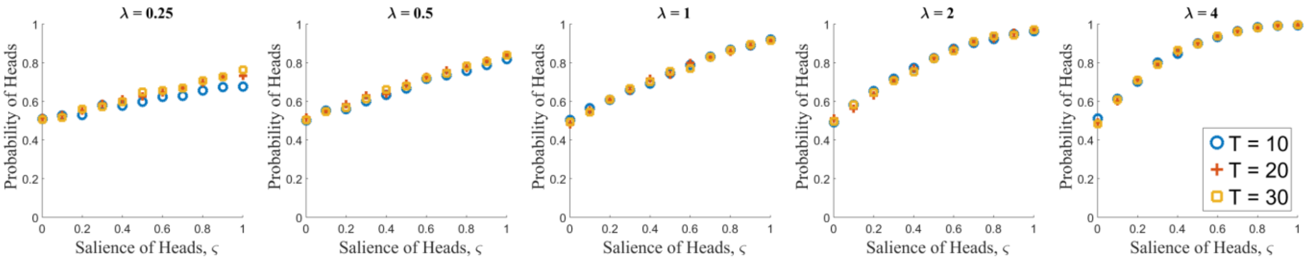

Figure 4: Simulated probability of heads in the simple heads-or-tails coordination game.

Salient Labels. Our model recognizes salience effects, too. In the simple heads-or-tails coordination game with heads being especially salient, such that $\sigma_{1\mathrm{H}} = \sigma_{2\mathrm{H}} = \varsigma$ and $\sigma_{1\mathrm{T}} = \sigma_{2\mathrm{T}} = 0$, we find that the probability of choosing heads is increasing in its salience ς, as shown in Figure 4. This figure plots the probability of choosing heads in this game as a function of the salience of heads, ς, for varying values of T and λ. As we should intuitively expect, when sampling is less noisy, i.e., when λ is greater, the players are more sensitive to salience. Specifically, when near the high end of our range, i.e., $\lambda = 4$, if heads is sufficiently salient, it is almost certain to be chosen. (In contrast, with an assumption of uniformly random sampling, our model would not account for any salience effect at all.) Convergence occurs quickly, so we see few effects from increasing the decision time T. Higher values of T only slightly reduce noise and increase the choice probability of heads when the logit sampling parameter λ is small.

Discussion

We have proposed a cognitive model of strategic deliberation and decision making. Our model is able to account for violations of Nash equilibrium involving failures of unravelling, payoff sensitivity, predictable coordination, and salience, and we illustrate this by simulating our model on four different games. Note that these violations have also been documented in a number of additional games, including the minimum-effort coordination game, the stag hunt game, the battle of sexes game, the discoordination game, the 11-20 game, the hide and seek game, the matching pennies game, and the Kreps game. Elsewhere we show that our model makes realistic behavioral predictions for all of these games, for $\lambda \in [0.25, 4]$ and $T \in [10, 30]$, however we exclude these findings from this paper, due to space constraints.

Our model is closely related to existing accumulator theories choice, and we suggest that it can be seen as a direct extension of decision field theory (Busemeyer & Townsend, 1993; also see Busemeyer, 2015 for a review). The novel element in our model involves the representation of beliefs regarding opponent's choices and the bidirectional updating of both preferences and beliefs over the time course of the decision process. Intuitively, bidirectional feedback in the accumulation process allows decision makers to base their choices on their beliefs about the opponent's choices, but also to update their beliefs as their own preferences evolve. As this updating happens gradually over time, the decision makers' intended choices (and beliefs about the opponent's choices) get increasingly more sophisticated the longer they spend deliberating. Eventually the nodes for the opponent's strategies develop unequal activation, with strategies that are appropriate responses to the decision maker's preferences having higher activation. Highly activated opponent strategies are more likely to be sampled, and the decision maker is subsequently more likely to develop preferences that intelligently respond to the opponent's anticipated choices.

Note that there is considerable evidence that decision makers are able to represent the preferences and beliefs of others separately from their own. Although the nature of these representations is not typically studied in the context of game theoretic deliberation, some experimental work on theory of mind in strategic games does support our proposed model. Hedden and Zhang (2002), for example, find that players in sequential move games have sophisticated beliefs about the opponent's preferences, and that these beliefs are dynamically modified based on the evidence presented to the decision maker during the decision process. Goodie et al. (2012) also find that players' beliefs about their opponent's preferences are fairly complex, and are formed in response to the players' own preferences.

Our approach is also closely related to cognitive decision modeling (in non-strategic settings) that uses neural networks with recurrent connectivity (Glöckner et al., 2014; Holyoak & Simon, 1999). Recurrence in these networks is often bidirectional; the activation of cues and decision attributes may influence and be influenced by beliefs and preferences. The bidirectional feedback in the above models and in ours is very similar, implying that our model could be adapted for other cognitive decision modeling applications.

Our bidirectional accumulation model also bears some resemblance to models of behavioral game theory, such as level-k reasoning and logit quantal response equilibrium (McKelvey & Palfrey, 1995; Nagel, 1995). In both our model and in level-k reasoning, individuals engage in an iterative process of deliberation that terminates before reaching a point of self-consistency. Likewise, in both our model and in logit quantal response equilibrium, individuals use a stochastic logit response rule to consider responses, thereby generating payoff sensitivity. However, unlike these models, our approach implements the deliberation process within a well-established psychological framework. This allows our model to describe salience effects, while also predicting the effects of time pressure and response time. Our model also makes more realistic stochastic choice predictions than either of these two existing theories: It permits trial-to-trial variability in choice, while also avoiding the selection of dominated strategies.

Our approach is also quite parsimonious. There are two parameters in our model: the decision time parameter, T, and the stochastic sampling parameter, λ. Decision time T can be seen as controlling the extent of bidirectional processing one can engage in during deliberation and thus determining one's level of strategic sophistication. Quick decisions involve fairly limited reasoning, with choices responding to simplistic beliefs about the opponent. Decisions that are a product of extended deliberation, in contrast, generate choices based on a more sophisticated theory of mind. As in all accumulator models, decision time also influences the amount of variability in the decision.

The stochastic sampling parameter λ can also be seen as affecting the extent of bidirectional processing one engages in. When λ is small, strategies are sampled with close to uniform probability, and activation in one layer of the network has little or no effect on the accumulation of activation in the other layer of the network. As λ increases, the decision maker becomes more and more likely to sample the most preferred strategies. When λ is very large, the most highly activated strategies are almost deterministically sampled, so preferences and beliefs interact more strongly during the deliberation.

Ultimately, the model's key behavioral properties depend critically on its dynamic and stochastic processes. Many scholars have suggested that behavioral theories of decision making can, with incorporation of these fundamental cognitive processes, describe a wide range of behavior (e.g. Busemeyer, 2015). Our results reinforce these claims by demonstrating the explanatory power of stochastic sampling and dynamic accumulation in strategic choice.

References

Basu, K.. (1994). The Traveler's Dilemma: Paradoxes of Rationality in Game Theory. *American Economic Review*, 84(2), 391-395.

Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin and Review*, 21(5), 1095-1111.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432-448.

Busemeyer, J. R. (2015). Cognitive science contributions to decision science. *Cognition*, 135, 43-46.

Camerer, C.(2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, 26(2), 139-198.

Glöckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition*, 133(3), 641-666.

Goodie, A. S., Doshi, P., & Young, D. L. (2012). Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, 25(1), 95-108.

Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, 85(1), 1-36.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, 128(1), 3-18.

Luce, R. D., & Raiffa, H. (1957). *Games and Decisions*. New York: John Wiley Sons.

Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, 84(3), 658-673.

Misyak, J. B., & Chater, N. (2014). Virtual bargaining: a theory of social decision-making. *Philosophical Transactions of the Royal Society: B*, 369(1655)..

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, 85(5), 1313-1326.

Rieskamp, J. (2006). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6), 1355.

Rubinstein, A. (2007). Instinctive and cognitive reasoning: a study of response times. *The Economic Journal*, 117(523), 1243-1259.