A Three-Pronged Simonesque Approach to Modeling and Simulation in Deviant "Bi-Pay" Auctions, and Beyond*

Joe Johnson[1] • Naveen Sundar G.[1] • Selmer Bringsjord[1,2,3]

Department of Computer Science[1]
Department of Cognitive Science[2]
Lally School of Management & Technology[3]
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA

Wednesday 12$^{\text{th}}$ March, 2014 16:59

Abstract. In order to employ and exhibit our Simon-inspired approach to computational economics, and specifically defend our version of the view that even logically untrained humans are rational, albeit no more than "boundedly" so, we provide two models, both rooted in computational logic, of how it is that logically untrained humans perform in a seemingly irrational fashion in a particular "deviant" auction (the bi-pay auction).

* We are profoundly indebted to the editor for wise guidance, and to two anonymous referees for insightful comments. We express gratitude too to Arturo Estrella, Ken Simons, and Will Tracy for suggestive comments on an earlier draft of this paper.

Table of Contents

List of Figures

1 Introduction

Bounded rationality concerns itself with how humans make decisions when it's unlikely or impossible for them to strictly apply such ideal and formalized techniques as Bayesian maximization of a utility function. However, as Selten (1999) explains, bounded rationality is not *irr*ationality: it isn't decision-making compromised by mental impairment or insanity. Instead, boundedly rational decision-making is triggered by circumstances in which either the limited processing power of the human mind, or the specific context at hand, renders ideal techniques either beyond reach, or within-reach but unattractive. According to Simon (1972), there are three potential causes of bounded rationality, each arising from limited power: uncertainty of the alternatives' consequences; incomplete information; and unconquerable computational complexity. Ideal forms of rationality can be formalized, but can such departure from the ideal also be formalized?

Selten (1999) argues for a negative answer. But we answer in the affirmative, and are building formal models of varieties of bounded rationality via logic-based computational cognitive modeling (an approach described e.g. in Bringsjord 2008) to confirm our optimism. In what follows herein, we report on our attempt to tackle a particular challenge: viz., the logic-based modeling of boundedly rational human decision-making in what we call the **bi-pay auction**. For us, such a challenge is met when we achieve a class of machine-discovered and machine-verified theorems that informatively characterize this non-ideal cognition. In general, then, our work is deeply rooted, not only in Simon's conception of bounded rationality, but also in other seminal work of his, including work that is generally neither studied nor harnessed by economists (and in our experience, sometimes not even known by them), but that is revered by computational logicians, and indeed by many working in AI: his stunning computer program LOGIC THEORIST,[1] which stole the show at the conference that founded the field of AI. LOGIC THEORIST was able to discover and verify an astounding number of theorems that had hitherto been the province only of human beings, and it marked the dawn of the field of automated theorem proving, still steadily progressing today. Furthermore, since our work, as will be seen, is aimed at computational simulation, we see ourselves as aligning with Simon's seminal investigation of the simulation of organizations (March & Simon 1958).

In the bi-pay auction (NOVA 2010), the highest bidder for an offered item pays her bid for that item; nothing new here. But there's a catch: the second-highest bidder *also* must pay her bid to the auctioneer—for nothing in return. The NOVA program featured a noteworthy experiment in which the prize was a $20 bill. Assuming "Bayes-max" behavior, no one should bid over $20. (It's

[1] This program is discussed in more detail at

http://en.widipedia.org/wiki/Logic_Theorist#Citations

and in Russell & Norvig (2009) in connection with the original AI conference in 1956 at Dartmouth.

provable that if this form of ideal rationality reigns in all agents, no one enters the auction.) Yet, players did the opposite: not only did they enter, but they continued bidding well above the 20-dollar level. In fact, bidding continued until a "winner" was produced at 28 dollars, at which point the second-highest bidder was (painfully) forced to pay his bid of 27 dollars to the (happy) auctioneer.[2] Assuming the players in the bi-pay auction are not insane, how can this outcome be explained? Informally put, the change to the rules of a standard auction, although subtle, make for a deviant auction; this presents a profound change to the risk/reward structure of the game. Yet the players must decide upon their strategy with limited time and resources; hence they move from ideal to bounded rationality; and they end up making mental mistakes against their own interests.

This is just a story. How can it be formalized? In the present paper, we present two formal models of bounded rational decision-making in a bi-pay auction. In both models we employ the "deontic cognitive event calculus" $\mathcal{DCEC}^*$ (Bringsjord & Govindarajulu 2013, Arkoudas & Bringsjord 2009) which provides a robust logico-computational framework for modeling multiple interacting agents, and their knowledge, beliefs, etc.

Our first formal model assumes that the boundedly rational player perceives a similarity between the bi-pay auction and a traditional auction. We mentioned the wrinkle about the second-highest bidder having to pay. Should that really deeply affect the way the game ought to be played? In fact, the difference is crucial, and the failure of a player to recognize that it is, is at the core of the first model's theorem that the observed behavior is to be expected when subjects rapidly "import" the formal structure of a standard auction.

Our second model reflects (in Selten's (1999) terminology) an *analytical* approach that assumes the agent analyzes the problem structure, identifies alternatives, and on that basis envisages consequences based on the beliefs of players about the beliefs of players, and what actions will flow from these iterated beliefs. A number of theorems are obtained. If for example the player knows that no other player will bid any more than \$5, the boundedly rational decision is provably to play. But it can also be proved that if the player merely believes this with high confidence, playing is boundedly rational, and consistent with a poker-player strategy; this explains why in some instances of the bi-pay auction bidding does stop below \$20.

2 Methodology

Though we confess that the phrase has a rather pedantic air about it, our methodology for building the models presented below can best be labeled **logic-based computational cognitive economic modeling**. As we have already

[2] This behavior has been replicated time and time again, with bidding for a prize not merely of \$20, but of \$100, with losers having to fork out as much as \$500! We have replicated the behavior ourselves, most recently on September 12 2013, when Bringsjord auctioned off \$100 for two bids of \$103 and \$102.

pointed out, historically speaking, this approach is rooted in three different research trajectories inaugurated by Simon: bounded rationality; cognitive modeling and simulation of firms; and the ability of computing machines to automatically prove substantive theorems. Viewed not historically, but technically, the foundation of this approach is the expression, in a formal language, of the elements of the given problem domain that give rise to rational action: knowledge, belief, perception, intention, and communication. Formal languages that include such cognitive elements allow us to leverage the inference rules of the language, in order to establish a computational framework in which we can model the (boundedly) rational actions of agents, and the justifications for these actions. Given the immense computing power current digital technology provides, and the expressivity of recently developed formal languages, this approach, we believe, shows impressive potential as a tool for studying human interaction, and by extension, economics. Whereas traditional models are formal, they often lack the granularity to sufficiently model the nuances of human behavior in the real world. Our approach, as shown below, can be applied at arbitrary levels of granularity; this in turn enables the capture of the full spectrum of factors driving rational action.

3 The Bi-Pay Auction

A bi-pay auction, as explained in the introduction, is an auction very similar to a normal auction—but with one subtle wrinkle: the second-highest bidder must pay his bid, and receives nothing in return. Though subtle, this change has profound consequences: it has been observed repeatedly in such auctions that participants usually end up bidding higher than the value of the prize being auctioned off. Because of this seemingly irrational behavior, we target the bi-pay auction with our Simonesque methodology.

4 Simulations

In this section we give first a detailed description of the computational model of the behavior of bidders in a normal auction, and then, second, in a bi-pay auction; in both cases we use $\mathcal{DCEC}^*$. As indicated above, $\mathcal{DCEC}^*$ provides a framework for modeling multi-agent systems wherein agent-to-agent interactions arise from the knowledge, beliefs, perceptions, plans, and natural-language capacity of the agents in question. The syntax and inference rules of the $\mathcal{DCEC}^*$ are described in detail in (Bringsjord & Govindarajulu 2013, Arkoudas & Bringsjord 2009); this machinery will be used heavily here. We specifically use the version of $\mathcal{DCEC}^*$ described in (Arkoudas & Bringsjord 2009).

The modeling is done in the Slate proof-construction environment. Slate combines both proof verification and proof discovery in an easy-to-use environment. Slate is a graphical proof-construction environment generally based on natural deduction, and includes support for constructing proofs in propositional logic, first-order logic, and several modal logics. Slate also has the ability to automatically discover proofs via resolution, by calling ATPs; for example, SNARK (Stickel 2008). This feature allows one to utilize Slate in a hybrid mode to construct proofs that are semi-automated. For an overview of an earlier version of Slate, see (Bringsjord, Taylor, Shilliday, Clark & Arkoudas 2008). Slate is ultimately based on a mathematical model of computation and reasoning that is a generalization and extension of Kolmogorov-Uspenskii machines (Kolmogorov & Uspenskii 1958, Bringsjord & Govindarajulu 2011).

A large portion of our subsequent discussion will focus on a simulation in which we formally model the activity in a normal auction between exactly two bidders in the first time interval, from time $t = 0$ to time $t = 1$, when the first bid is placed. (Constraining our model to two bidders limits the number of formal expressions to a manageable size for exposition here. However, the same approach can be extended to apply to the general case of an auction with n bidders.) This simulation will provide the reader with an understanding of how the model works, convey some of its power, and specifically show how this approach can serve as a useful tool in economic modeling. Finally, we employ the model to prove two critical assertions about the two auctions: one, in the normal auction, bidding does not exceed \$20; and two, in the bi-pay auction, it does.

4.1 Normal Auction

Normal Auction Axioms We begin with a set of "auction axioms" that formally express the rules of a normal auction.[3] Our first axiom for the normal auction is: It is common knowledge that at all points in time during the game, $t > 0$, for all players, a, and for all bid amounts $x > 0$: player a bids $x + 1$ at time $t + 1$ only if at time t, x is the high bid and player a is not the current high bidder.

[3] In light of the likely event that our readers are familiar with so-called "axiomatic approaches" in economics, we here point out that our use of the term 'axiom' is fundamentally different than the use to which those in e.g. the "Nash program" have put it. (We mention for purposes of quick exposition only one line of work under the axiomatic approach.) Our use of the term 'axiom' is tied directly to terminological custom in AI, and to Simon's founding of that field and custom via providing (in the aforementioned year of 1956) *formal* axioms (from *Principia Mathematica*) over which automated theorem proving can (advantageously) happen. In contrast, the axiomatic approach of Nash and those in his "program" (see e.g. Nash 1953, Binmore, Rubinstein & Wolinsky 1986) revolves around axioms that partake in no small part of natural language (e.g., English), which makes these "axioms" unsuitable for machine processing, but ideal for human consumption, understanding, and progress. A parallel point could be made in connection with the "axioms" of game theory, which are expressed partly in English, and partly in formal language. Our Simonesque program, longer term, is to formalize the axioms in play in axiomatic approaches in microeconomics so as to enable intelligent computing machines to participate in discovering novel consequences and application of these axioms.

$$[AA1]\ \mathbf{C}(\forall t, a, x.\ happens(action(a, bids(x+1), t+1)) \\ \Rightarrow holds(highbid(x), t) \wedge \neg holds(highbidder(a), t))$$

For our next axiom: It is common knowledge that at all times, t, for all players a and for all bid amounts x: player a is the high bidder at time t if and only if player a successfully bids x at time t.

$$[AA2]\ \mathbf{C}(\forall t, a, x.\ holds(highbidder(a), t) \Leftrightarrow happens(action(a, bids(x)), t))$$

Similarly, it is also common knowledge that x is the high bid at time t if and only if player a successfully bids x at time t.

$$[AA3]\ \mathbf{C}(\forall t, a, x.\ holds(highbid(x), t) \Leftrightarrow happens(action(a, bids(x)), t))$$

Next, it is common knowledge that there is no more than one high bidder at any time t.

$$[AA4]\ \mathbf{C}(\forall t, a, y.\ (holds(highbidder(a), t) \wedge y \neq a) \Rightarrow \neg holds(highbidder(y), t))$$

It is also common knowledge that there is no more than one high bid at any time t.

$$[AA5]\ \mathbf{C}(\forall t, a, x.\ (holds(highbid(x), t) \wedge y \neq x) \Rightarrow \neg holds(highbid(y), t))$$

At all times t, if an agent makes a bid, then all agents perceive it. That is, this is a public auction where all bidding activity is known to everyone.

$$[AA6]\ \forall t, a, x, b.\ happens(action(a, bids(x)), t) \Rightarrow \mathbf{P}(b, happens(action(a, bids(x)), t))$$

It is common knowledge that at all times t: bidder a buys item at time $t+1$ for price x if and only if bidder a successfully bids x at time t, and there does not exist a bidder b, $(b \neq a)$, who bids $x+1$ at time $t+1$.

$$[AA7]\ \mathbf{C}\forall t, a, x.\ happens(action(a, buys(item, x)), t+1) \\ \Leftrightarrow (happens(action(a, bids(x)), t) \\ \wedge \neg\exists b.\ (b \neq a \wedge happens(action(b, bids(x+1)), t+1))$$

Normal Auction Agent Beliefs, Desires, and Intentions The following expresses the beliefs of each agent in the auction about the value of the item up for bid: the $20 bill. Unlike other auctions where the items for bid may be valued differently based on subjective judgements of the players involved, in this auction, the value of the $20 bill is plainly obvious to everyone.

$$[B1]\ \forall a.\ \mathbf{B}(a, \mathit{rewardvalue}(20))$$

Finally, the following formula gives the strategy employed by the bidders (for the 2-player auction): Each bidder places a bid if he believes the value of the item is $20, if he knows the high bid is x and $x < 20$, and if he is not currently the high bidder.

$$\begin{aligned}[I1]\ \forall t, a, x.\ & \mathbf{B}(a, \mathit{rewardvalue}(20)) \\ & \wedge \mathbf{K}(a, \mathit{holds}(\mathit{highbid}(x), t)) \\ & \wedge \mathbf{K}(a, \neg \mathit{holds}(\mathit{highbidder}(a), t)) \\ & \wedge \mathbf{K}(a, (x < 20)) \\ & \Rightarrow \mathit{happens}(\mathit{action}(a, \mathit{bids}((x+1), t+1)\end{aligned}$$

Normal Auction Proof Using the formulae established above for the axioms of the auction and the strategy employed by the bidding agents, we now construct a formal proof that models the bidding behavior and interactions of the agents, along with inherent justification for each of the bidding events. Finally, we present a proof showing that there is no bidding above $20. It is important to note that proof-discovery and proof-checking are the very means by which we achieve simulation. In this regard, it is not inaccurate to say that we weave together two seminal Simonesque strands (modeling and simulation, and automated theorem proving) that Simon himself, to our knowledge, never integrated. In addition, unlike cases where modeling and simple simulation are achieved by *informal* proofs, we in one fell swoop achieve formal verification of our processing by proof checking. Today's advances in automated theorem proving make this possible, which puts to rest an awkward period during which machine reasoning was *believed* to be valid, but not *known* to be valid (for a fuller discussion, see e.g. Arkoudas & Bringsjord 2007).

We start with some formal expressions concerning the state-of-affairs prior to the start of bidding. Let's assume the names of our agents are $A1$ and $A2$. We make the following claims with respect to agent $A1$: First, everyone knows the high bid is $0 before bidding begins.

$$[D1]\ \forall a.\ \mathbf{K}(a, \mathit{holds}(\mathit{highbid}(0), 0))$$

Next, everyone knows that $A1$ is not the current high bidder. For the purposes of technical convenience, we'll assume that everyone regards $A2$ as the high bidder at $t = 0$, even though that is not the case, of course. However, this assumption makes it easier to see how the auction proceeds in a back-and-forth pattern between $A1$ and $A2$ over time, without, from a practical perspective, reducing the overall veracity of the model.

$$[D2]\ \forall a.\ \mathbf{K}(a, \neg holds(highbidder(A1), 0))$$

$$[D3]\ \forall a.\ \mathbf{K}(a, holds(highbidder(A2), 0))$$

Finally, we assert that it's common knowledge that $A1$ and $A2$ refer to different bidders.

$$[D4]\ \mathbf{C}(A1 \neq A2)$$

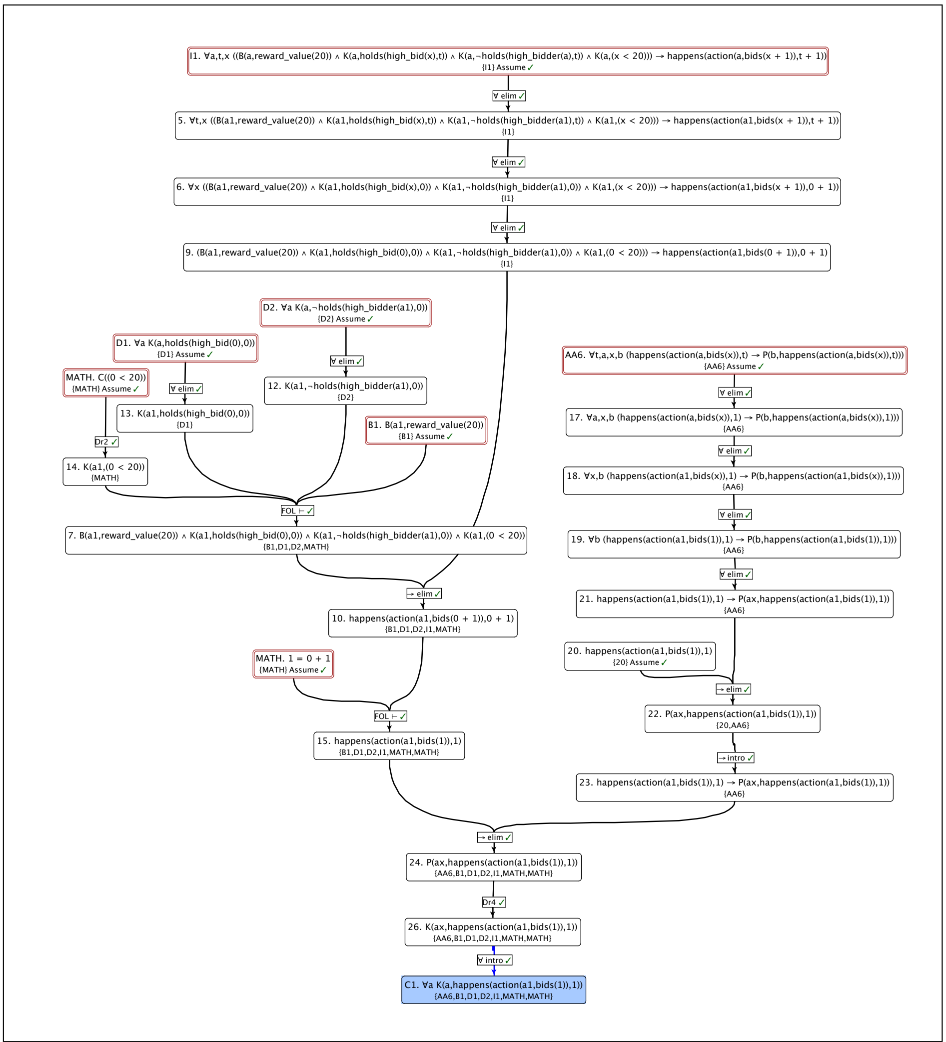
These declarations, along with the axioms, beliefs, and intentions of the agents, are sufficient to model the bidding behavior of the agents in the auction. The formulae below, which express key developments at the start of the auction from time $t = 0$ to time $t = 1$ (after the first bid is placed), serve multiple purposes. First, they show that at each step of the auction, we can not only model behavior, but provide a justification of the events as they occur. Second, proofs of these assertions demonstrate the veracity of the model. And third, and perhaps most imporantly, they show that this approach can be used as a form of simulation in which the formal expression of knowledge, beliefs, desires, and intentions serve as the basis for computational modeling of (boundedly) rational interactions among multiple agents.

The first conclusion we can infer is that every bidder knows that $A1$ bids \$1 at time $t = 1$, as shown in Figure 1. This inference can be drawn by applying the inference rules of the $\mathcal{CEC}^*$ and $\mathcal{FOL}$ (first-order logic) to the premises, $[AA6]$, $[B1]$, $[I1]$, $[D1]$, $[D2]$, and a pair of trivial axioms of arithmetic, as shown in Figure 1.

$$[C1]\ \forall a.\ \mathbf{K}(a, happens(action(bids(A1, 1)), 1))$$

Next, we can infer our second conclusion, that every bidder knows that $A1$ is the high bidder at time $t = 1$, as shown in Figure 2. This inference can be drawn from $[C1]$ and $[AA2]$.

Fig. 1. Proof of $C1$. The conclusion that $A1$ bids \$1 at time $t = 1$ is formally proved using $[AA6]$, $[B1]$, $[I1]$, $[D1]$, and $[D2]$.

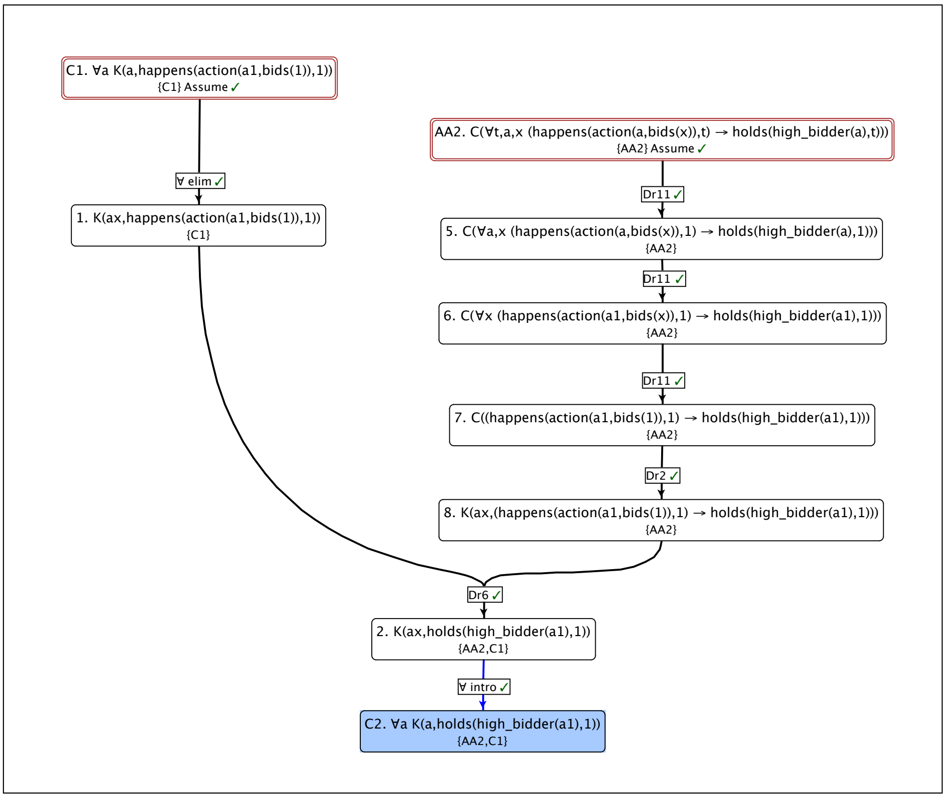


$$[C2]\ \forall a.\ \mathbf{K}(a, holds(highbidder(A1), 1))$$

Using $[C1]$ (again) and $[AA3]$ we can infer our third conclusion, that every bidder knows the high bid is \$1 at time $t = 1$. This proof is shown in Figure 3.

$$[C3]\ \forall a.\ \mathbf{K}(a, holds(highbid(1), 1)$$

Fig. 2. Proof of $C2$**.** We infer $A1$ is the high bidder at time $t = 1$ from $[C1]$ and $[AA2]$.



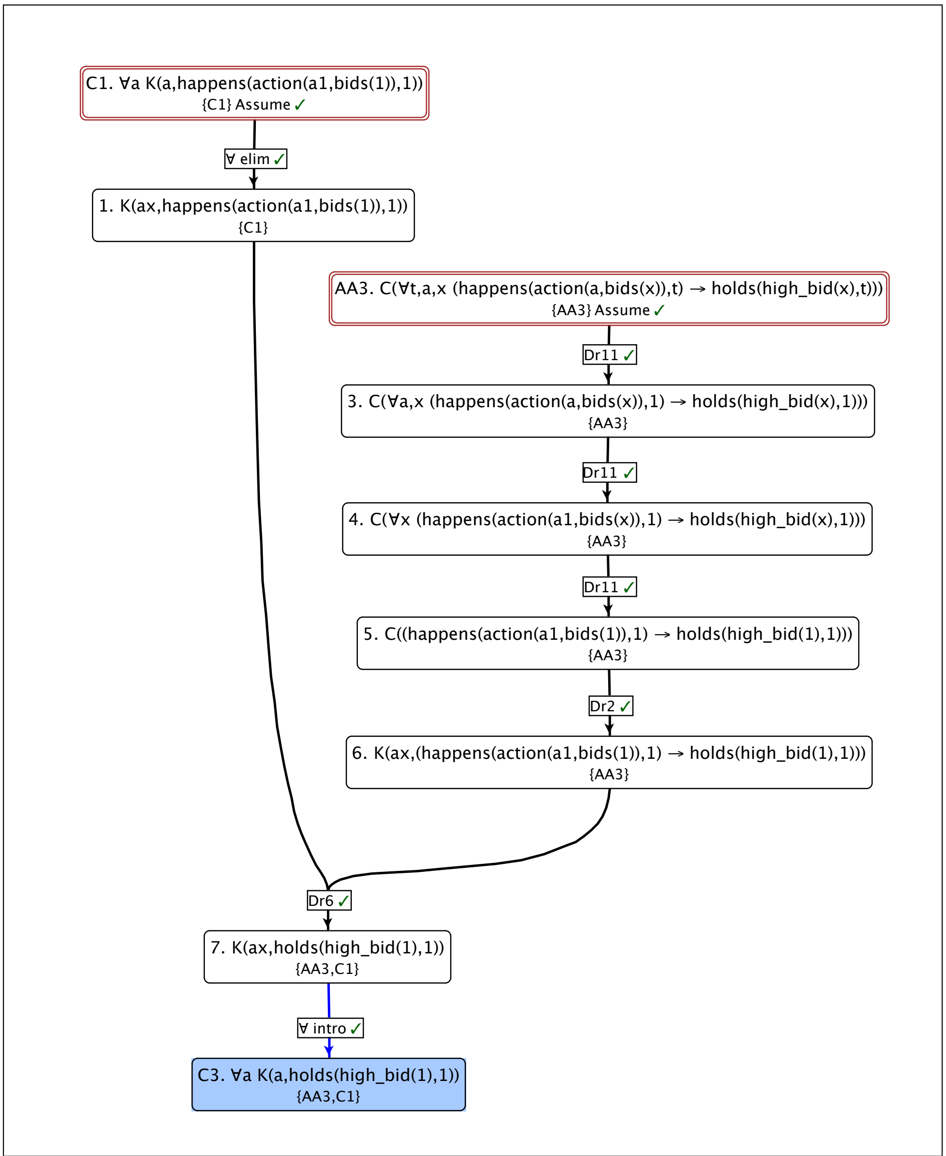
Next, we can infer that every bidder knows it does not hold that $A2$ is the high bidder at time $t = 1$, as shown in Figure 4, using $[C2]$, $[AA4]$, and $[D4]$.

$$[C4]\ \forall a.\ \mathbf{K}(a, \neg holds(highbidder(A2), 1))$$

Notice that with $[C2]$, $[C3]$, and $[C4]$ in hand, we have inferred the same information about the state of the auction at time $t = 1$ as we had asserted at time $t = 0$, namely, that everyone knows what the high bid is, who the current high bidder is, and who the current high bidder is not. From here, we could continue in the same manner to infer that everyone knows the same about the state-of-affairs at time $t = 2$, and, in turn, at $t = 3$, and so on.

As a further demonstration of the model, however, we show a couple more inferences that can be drawn about the state at time $t = 1$. First, we can infer that every bidder knows it does not hold that \$0 is the high bid at time $t = 1$,

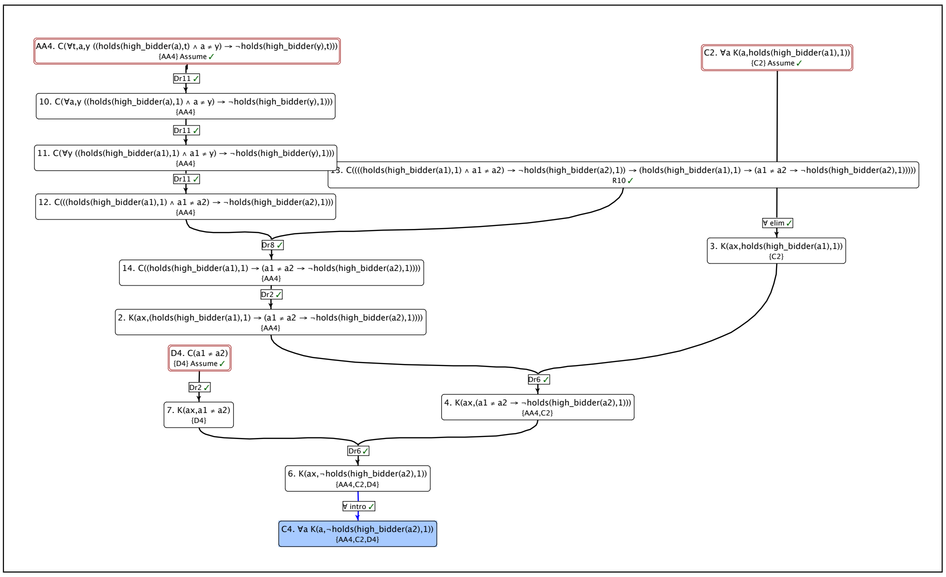
Fig. 3. Proof of $C3$**.** The conclusion that every bidder knows the high bid is \$1 at time $t = 1$ is inferred from $C1$ and $AA3$.



the proof for which is in Figure 5, utilizing [$AA5$], [$C3$], and a trivial fact of arithmetic.

$$[C5]\ \forall a.\ \mathbf{K}(a, \neg holds(highbid(0), 1))$$

Fig. 4. Proof of $C4$. The inference that every bidder knows it does not hold that $A2$ is the high bidder at time $t = 1$ is drawn from $[C2]$, $[AA4]$, and $[D4]$.



And second, we infer that $A2$ does *not* bid \$1 at time $t = 1$, as shown in Figure 6 using $[AA1]$, $[D3]$, and arithmetic.

$$[C6]\ \neg happens(action(A2, bids(1)), 1)$$

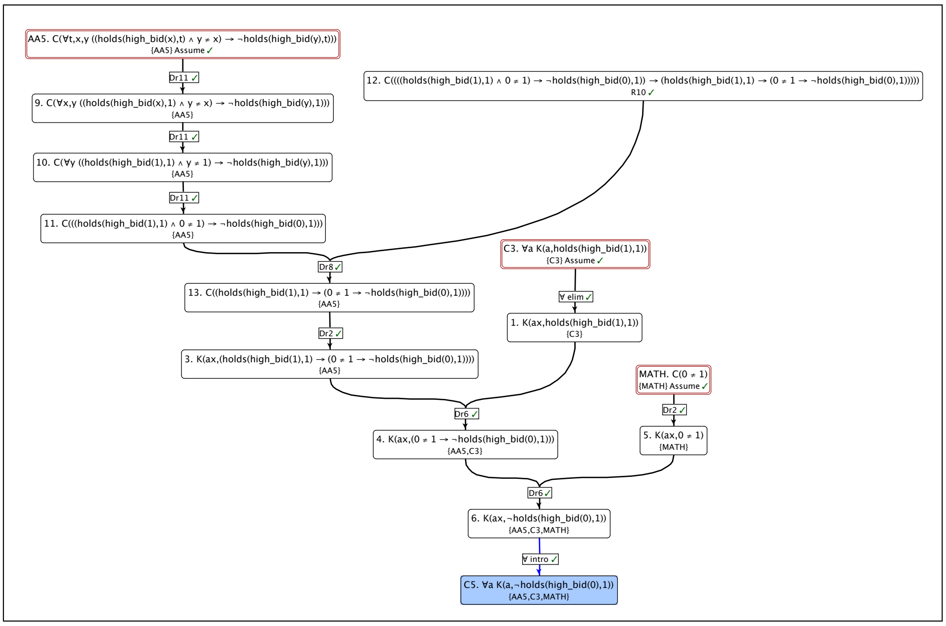
Lastly, we prove something important about this "normal" auction: that bidding never exceeds \$20. (In the next section, we will formally show exactly the opposite for the bi-pay auction: that bidding will proceed above \$20.) We start by proving the following intermediate result, as shown in Figure 7, that if someone bids $x + 1$ at time $t + 1$, then $x + 1 \leq 20$:

$$[IR1]\ \forall a, t, x.\ (happens(action(a, bids(x + 1)), t + 1) \rightarrow (x + 1 \leq 20))$$

Using $[IR1]$, we can go on to prove the following formula, in Figure 8, which formally re-states $[IR1]$ in a more direct way.

$$[C7]\ \neg\exists x, a, t.\ (happens(action(bids(a, x + 1)), t + 1) \wedge x > 20$$

Fig. 5. Proof of $C5$. We formally prove every bidder knows $0 is not the high bid at time $t = 1$ from $[AA5]$, $[C3]$, and some assertions of arithmetic.



4.2 Bi-Pay Auction Model 1 (BR1)

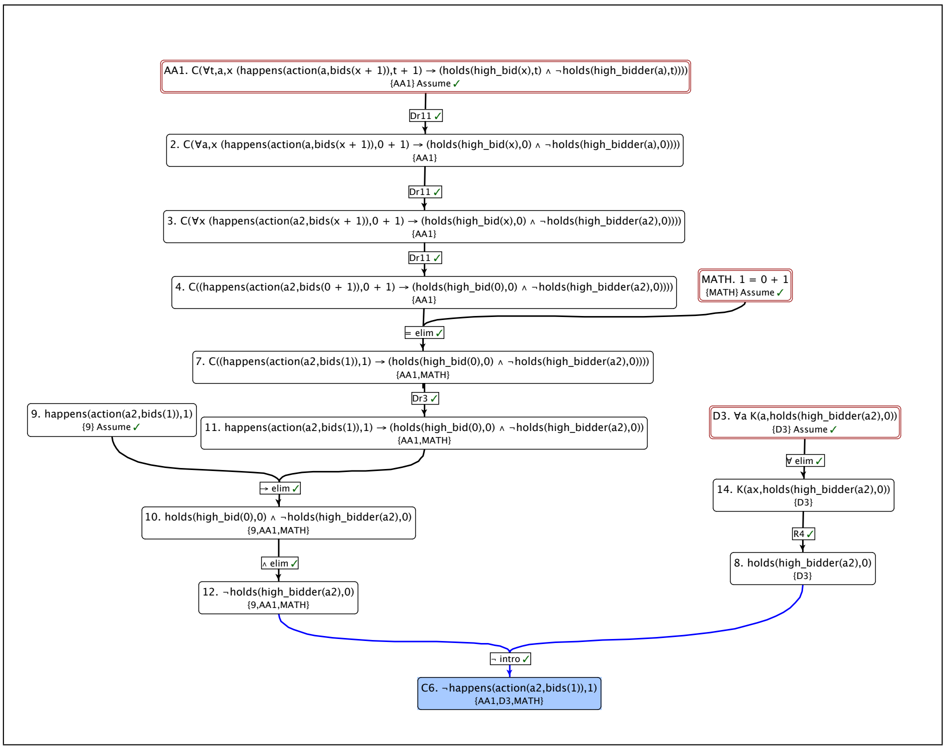
Bi-Pay Auction Axioms We move on, now, to discuss modeling the subtly more complex bi-pay auction, which, as we described, works the same way as a "normal" auction, except that the second-highest bidder must pay his or her bid (but receives nothing in return). In order to model an auction with this extra wrinkle, we must add a few more axioms.

First, it is common knowledge that for all t and all agents a: a is second-highest bidder at time $t+1$ if and only if a was high bidder at time t, and there exists bidder, b, who bids $x+1$ at $t+1$ ($b \neq a$). Formally, this can be expressed as follows:

$$
\begin{aligned}
[AA8]\ \mathbf{C}(\forall t, a.\ & holds(secondhighbidder(a), t+1) \\
& \Leftrightarrow (holds(highbidder(a), t) \\
& \wedge \exists b.\ happens(action(bids(b, x+1)), t+1) \wedge b \neq a)
\end{aligned}
$$

Next, it is common knowledge that for all t and bid amounts x, x is a second high bid at time $t+1$ if and only if x was high bid at time t and there exists a bidder, a, that bids $x+1$ at $t+1$.

Fig. 6. Proof of $C6$. The inference that $A2$ does *not* bid \$1 at time $t = 1$ is drawn from $[AA1]$, $[D3]$, and assertions of arithmetic.



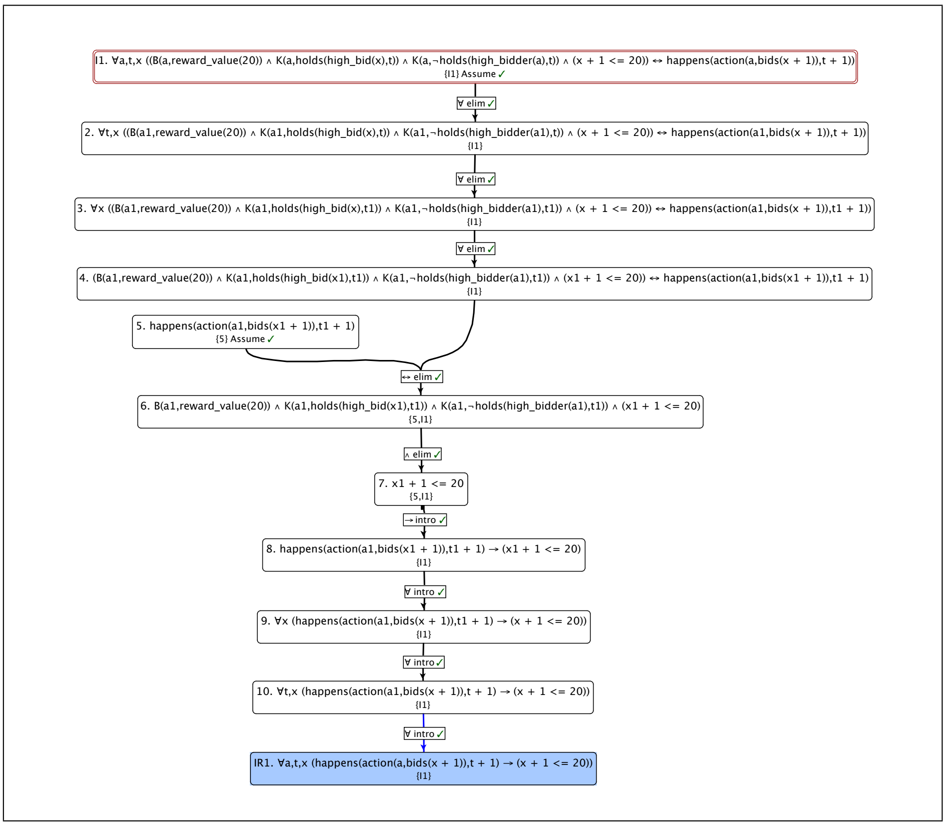
$$[AA9]\ \mathbf{C}(\forall t, x.\ holds(secondhighbid(x), t+1)$$
$$\Leftrightarrow (holds(highbid(x), t) \land \exists a.happens(action(bids(a, x+1)), t+1))$$

Last, it is common knowledge that for all times t, all agents a, and all bids x: a pays x at $t+1$ if and only if a is second-highest bidder at t, and there is not an agent b ($b \neq a$) who bids $x+1$ at $t+1$.

$$[AA10]\ \mathbf{C}(\forall t, a, x.\ happens(action(a, pays(x)), t+2)$$
$$\Leftrightarrow (holds(secondhighbidder(a), t+1)$$
$$\land \neg\exists b.\ (happens(action(b, bids(x+2)), t+2)))$$

Bi-Pay Auction Agent Beliefs, Desires, and Intentions In terms of agent beliefs in the bi-pay auction, there is an important belief that arises regarding

Fig. 7. Proof of $IR1$. This intermediate result, the conclusion that if someone bids $x + 1$ at time $t + 1$, then $x + 1 \leq 20$, is used in the proof of $C7$.

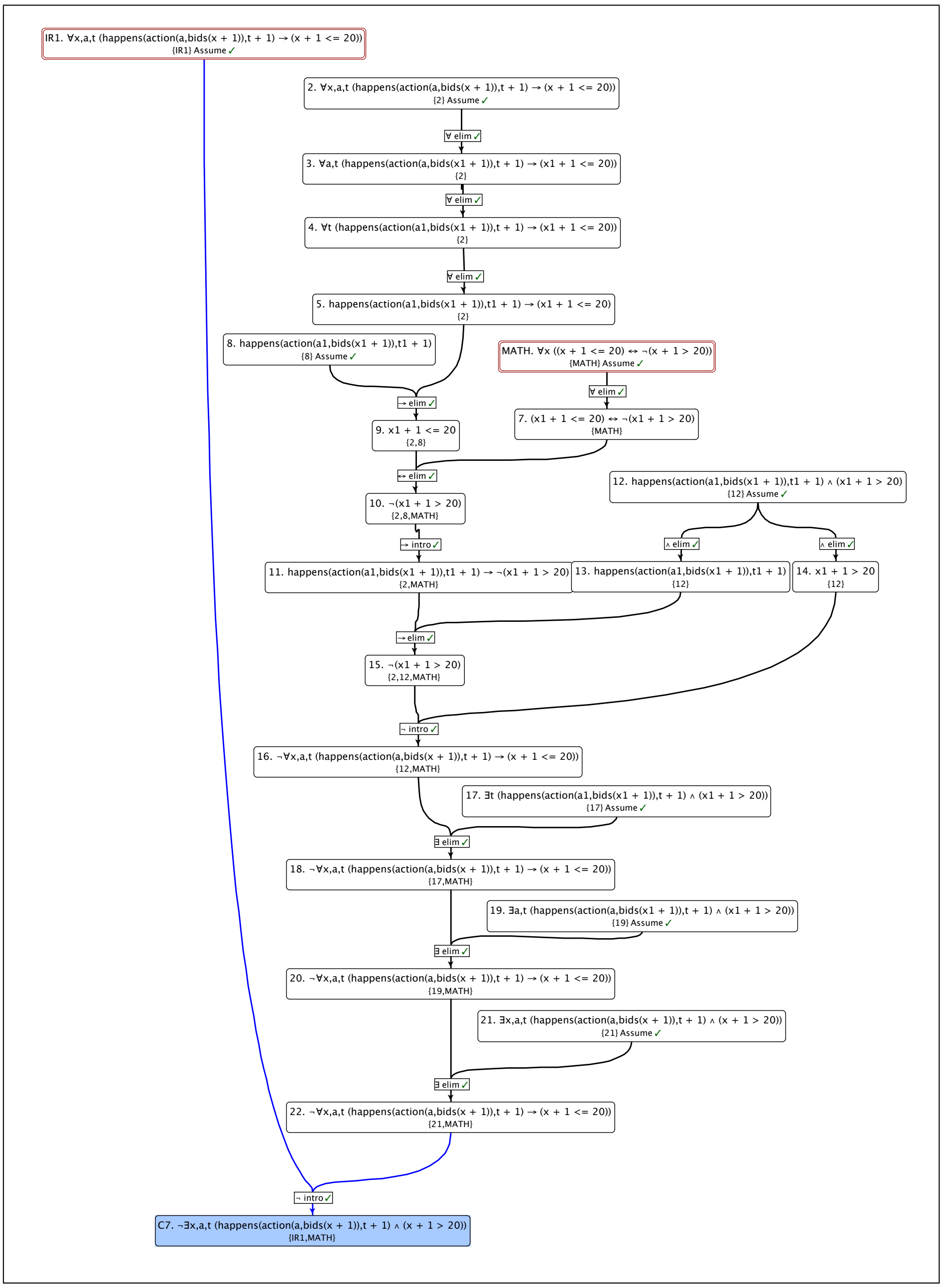


what every agent believes about the intentions of every other agent. This belief is relevant to the modeling of the seemingly irrational behavior shown in experiments on human subjects in bi-pay auctions. That is, in the rush to decide whether and how to bid in the bi-pay auction, the boundedly rational individual *imports* the knowledge and beliefs he possesses about a normal auction in order to make a quick decision about whether to play and, at least to a slight degree, how to play in this new sort of auction. One of those beliefs stems from what we observed in the normal auction: that no one bids more than the prize is worth (in this case, \$20). Formally, we can express this as follows:

$$[B2]\ \forall x, a, t, b.\ (a \neq b \rightarrow \mathbf{B}(a, \mathit{happens}(\mathit{action}(\mathit{bid}(x+1)), t+1) \rightarrow (x+1 \leq 20))$$

At some point, particularly as bidding approaches \$20, one, and eventually both, of the agents will recognize the precariousness of the second-highest bidder

Fig. 8. Proof of $C7$. Using $[IR1]$, we prove bidding does not go above $20.



position that didn't exist in the normal auction. (We have observed this recently in our own experiments based on not a \$20 but a \$100 "prize." Indeed, most recently, we heard a number of "Uh ohs.") It is for this reason that when the second-highest bidder is at, say, \$19, he may modify his strategy from $[I1]$ to one based on the following idea: If he's second-highest bidder and he doesn't believe the other player will bid $x+2$, then he'll go ahead and bid $x+1$ (which at this point in the auction translates to a bid of \$21—more than the prize amount). More formally, for all t, agents a, agents b ($b \neq a$), and bids x: if, at time t, a is second-highest bidder and high bid is x and a does not believe b will bid $x+2$ at $t+2$, then a will bid $x+1$ at $t+1$.

$$\begin{aligned}[I2]\ \forall x, a, t, b.\ &((\mathbf{K}(a, \mathit{holds}(\mathit{secondhighbidder}(a), t)) \wedge \mathbf{K}(a, \mathit{holds}(\mathit{highbid}(x), t)) \\ &\wedge \neg\mathbf{B}(a, \mathit{happens}(\mathit{action}(b, \mathit{bids}(x+2)), t+2)) \wedge b \neq a) \\ &\rightarrow \mathit{happens}(\mathit{action}(a, \mathit{bid}(x+1)), t+1))\end{aligned}$$

Lastly, for the sake of our proof in the next section, we need to formally assert that no bidding agent believes $22 \leq 20$, which is without question true.

$$[B3]\ \forall a.\ \neg\mathbf{B}(a, 22 <= 20)$$

Bi-Pay Auction Proof Based on these new beliefs and intentions, we formally model how bidding proceeds above \$20. We do this by focusing on time $t = 20$, in which $A1$ finds himself as second-highest bidder with a bid of \$19, and $A2$ as high bidder with \$20. $A1$ must make a difficult choice: bid \$21 or lose \$19. It's at this point (if not before) that he devises $[I2]$, to bid as long as he does not believe his opponent will outbid him. Since at this moment he carries the belief that no *other* agent would bid above \$20, he bids \$21. In order to formally infer this, we need to make some formal declarations about the state-of-affairs at $t = 20$. First, both bidders know that $A1$ is the second-highest bidder. And second, both bidders know that the high bid is \$20.

$$[D5]\ \forall a.\ \mathbf{K}(a, \mathit{holds}(\mathit{secondhighbidder}(A1), 20))$$

$$[D6]\ \forall a.\ \mathbf{K}(a, \mathit{holds}(\mathit{highbid}(20), 20))$$

Using $[D5]$ and $[D6]$ along with $[B2]$, $[B3]$, and $[I2]$, we deduce that there are bids that exceed \$20 during the bi-pay auction. The proof is shown in Figure 9.

$$[C8]\ \exists x, a, t.\ (\mathit{happens}(\mathit{action}(\mathit{bids}(a, x+1)), t+1) \wedge x > 20$$

4.3 Bi-Pay Auction Model 2 (BR2)

Finally, we show how bounded rationality could be modeled under the BR2 model. Under certain conditions, rational agents would perform actions which a shallow analysis of rationality might deem irrational. For example, in the bi-pay auction, everyone could have decided not to participate in the auction as they know the nature of this auction, but only one agent could know of this fact. A rational agent would make use of this fact and participate in this auction. Consider the following imaginary, but quite plausible, setup:

> In a class on "Human Irrationality" there are only two students: Alice and Bob. The teacher, Carol, invites Bob to her office and shows him a video demonstrating the conventional analysis of the bi-pay auction. The video smugly declares that it is irrational to participate in a bi-pay auction. The next day, the teacher invites Alice to her office and shows her the same video. When Alice is about to leave, the teacher offhandedly remarks that she had shown the same video to Bob. The teacher then meets Alice and Bob in the classroom the next day and launches a bi-pay auction for $100 to make sure that they have both learned the lesson the video was supposed to teach. The teacher hopes that both Alice and Bob remain silent. To her great surprise and anguish, Alice bids a dollar and wins $99 while Bob refrains from bidding.

Given the Simonesque machinery we have availed ourselves of, this behavior can be captured in a simple model $\mathcal{M}^{\mathsf{BR_2}}_{\mathsf{BiPay}}$ that accounts for agents that have iterative beliefs. The model captures that both Alice and Bob know the properties of the bi-pay auction. The model also captures, thanks to the teacher's offhand remark, that Alice knows that Bob knows about the bi-pay auction's general properties. Unfortunately, Bob does not know that Alice knows about the auction. Both Alice and Bob understand the video's dictum that if anyone does not know whether any other person knows about the bi-pay auction, they should not participate. This simple, but reasonable, model of rationality predicts that it is rational for Alice to bid and for Bob to refrain from bidding.

Generating the Prediction from a $\mathcal{DCEC}^*$ Model Departing from the previous two sections' intricate semi-automatic modeling, we generate the predictions from this model using *fully* automated reasoning. This is mainly to demonstrate that our models can not only be cognitively deeper, but also amenable to the present day's version of the automation that Simon introduced to the world via `logic theorist`. We use the multi-sorted first-order theorem prover SNARK (Stickel 2008); but any number of provers would suffice.

There are two general ways in which reasoning in a higher-order logic (either extensional, such as second-order logic, or intensional, such as modal logics) can be cast into first-order logic. In what we term "direct translation," one just represents syntactic rules in the proof calculus for the higher-order logic as a first-order theory that deals with symbols, formulae, inference rules etc.

This method, while accurate, is highly inefficient. In the second method, which can be coined "first-order approximation," we convert all the modal operators into first-order predicate symbols and create for all first-order predicate symbols P, logic connectives Θ ($\wedge$,$\vee$,...), corresponding function symbols P_f and Θ_f ($\wedge_f$,$\vee_f$,...). This method, while efficient, could result in inconsistencies when dealing with large amounts of sentences (Bringsjord & Govindarajulu 2012). We use the latter method below.

The first-order analogue for **K** and the mirrored function symbols can be seen in the listing below, which declares the signature of the multi-sorted first-order theory that we will be using. The listing is in Common Lisp and adheres to the interface provided by SNARK.

```

(defun decls ()
  (declare-sort 'Agent)
  (declare-sort 'Proposition)
  (declare-sort 'ActionType)
  (declare-variable '?a :sort 'ActionType)

  (declare-constant 'p :sort 'Agent)
  (declare-constant 'q :sort 'Agent)
  (declare-constant 'BIPAY :sort 'Proposition)

  (declare-constant 'refrain :sort 'ActionType)
  (declare-constant 'bid :sort 'ActionType)
  (declare-function 'action 2 :sort
              '(Proposition Agent ActionType))

  (declare-function 'knows-f 2 :sort
              '(Proposition agent Proposition))
  (declare-function 'not-f 1 :sort
             '(Proposition  Proposition))

  (declare-relation 'knows 2 :sort '(agent Proposition))
  (declare-relation 'Holds 1 :sort '(Proposition)))
```

Listing 1.1. Declaring Logic Sorts and Symbols

The second step is converting the first-order modal axioms into multi-sorted first-order axioms. The listing below shows one such conversion.

```
(forall ((?a :sort agent) (?P :sort Proposition))
     (implies  (knows ?a ?P) (Holds ?P)))
;;;; more such axioms and rules not shown
```

Listing 1.2. $\mathcal{DCEC}^*$ Axioms

The final step in the process is representing the model $\mathcal{M}^{\mathsf{BR_2}}_{\mathsf{BiPay}}$ as a set of first-order sentences. The listing below shows the axioms in the model. Axiom 1 asserts that we have only two agents p and q. Axiom 2 asserts that we have only two possible actions. Axioms 3 and 4 assert that both the agents know about the bi-pay auction. Axiom 5 asserts that p knows that q knows about the auction. Axioms 6 and 7 assert that q is ignorant as to whether p knows the same or not. Axiom 8 asserts that if any agent is unsure about whether some other agent knows about the auction, that agent should not participate in the auction.

```
;; Axiom 1:
 (forall ((?x :sort Agent))
     (and (not (= p q))
             (or (= ?x p) (= ?x q))))
;; Axiom 2:
   (forall ((?a :sort Agent))
     (iff (not (Holds (action ?a bid)))
          (Holds (action ?a refrain))))

;; Axioms 3 and 4
   (knows p BIPAY)
   (knows q BIPAY)

;; Axiom 5:
   (knows p (knows-f q BIPAY))

;; Axioms 6 and 7
   (not (knows q (knows-f p BIPAY)))
   (not (knows q (not-f (knows-f p BIPAY))))

;; Axiom 8
   (forall ((?a :sort Agent))
     (iff
     (exists ((?b :sort Agent))
        (and
         (not (= ?a ?b))
         (not (knows ?a (knows-f ?b BIPAY)))
         (not (knows ?a (not-f (knows-f ?b BIPAY))))))
     (Holds (action ?a refrain))))
```

Listing 1.3. Model Axioms

With these axioms in place, SNARK can prove the predictions given in the listing below almost instantaneously (around 10 milliseconds each on a commodity laptop).

```
(Holds (action p bid))
(Holds (action q refrain))
```

Listing 1.4. Predictions

5 Empirical Reflections

Because the research reported on herein, true to Simon's own *modus operandi*, spans a number of fields, the empirical dimension of the research ranges across the relevant fields.

If we think of our work in terms of experimentation within computer science, AI, and automated theorem proving, the upshot is that even though the modeling in question reaches down to the level of individual cognitive structures of individual agents, there is no reason to think that such a fine-grained approach will not allow modeling and simulation of a fully implemented sort, in reasonable time. We are of course not *per se* interested in modeling and simulating auctions, but rather, with Simon, organizations—but in light of our results there is no reason, empirically speaking, to think that the methodology here introduced and employed cannot scale to the size of even large corporations. By the time we apply our methodology to modeling a sizable organization, the hardware available to us will be significantly better, and we have yet to explore the parallelization of our methodology, which would of course enable us to implement our methodology on a supercomputer. While we do not know, we do at least ponder the possibility of macroeconomic simulation that reaches down to the level of individual belief, intention, and communication. At present, as the reader will doubtless know, macroeconomic computational simulation is done in the aggregate, far from the details of individual minds.

In a different sense of 'empirical,' it would seem fair to say that behavioral experiments match both models presented in this paper. Most such findings are consistent with our first model, BR1, as described in the introduction. In an experiment conducted recently by one of the authors of this paper, bidding in a bi-pay auction conducted in a class for a prize of \$100 continued up to \$103. (To be clear, all monies were returned to the participants after the experiment; but participants were unaware this would be the case until after the auction was over.) However, in a separate experiment, in a different class, where the prize was once again \$20 (not \$100), students refrained from bidding, seemingly aware of the pitfall of this auction. After some time, one student forcefully came forward with a bid for \$1. One other student meekly countered with a \$2 bid. When the first student immediately responded with a confident \$3 bid, bidding stopped, and the winning student walked away with \$17. This matches, or is at least consistent with, BR2.

6 Objections and Rebuttals

Objection 1: *Ad Hoc* Axioms?

The first objection we consider is as follows: "The model that explains the behavior in the bi-pay auction requires additional assumptions over and above the assumptions used to build the model of the normal auction. One of these additional assumptions is Axiom 8, which asserts that if agent_1 is unsure about whether some other agent_2 knows about the auction, agent_1 should not participate in the auction. Axiom 8 is peculiar to the bi-pay auction, and so isn't a *general* assumption about human behaviour — but such generality, surely, is what a psychologist would expect of a model of the behaviour of participants in the bi-pay auction. Presumably Axiom 8 should be able to be deduced by an agent (and so by a model of the agent) who analyzes the possible outcomes of participating, by using, for example, a decision tree."

This is an insightful objection. But we use the term 'axiom' (in this specific portion of our model) to denote a deep, foundational principle for a class of cognizers, and we explicitly countenance different such classes, each characterized by differential axiom sets. Theorems flow from axiom sets, and since decision trees are formally equivalent to (quantified) conditionals in first-order logic, that which a tree generates wouldn't for us be an axiom. Moreoever, one agent able to shift between axiom sets, and affirm one set as opposed to another in keeping with a given context, while an intoxicating prospect, is beyond the scope of our research and development at this point. At present, we accept that, with respect to reasoning, especially specifically deductive reasoning, there are vast individual differences between types of human (and, for that matter, machine) reasoners, and we rest content with formalization of each type. This appproach is quite consistent with formalization in other domains. For instance, even in the case of axiomatic set theory, uncontroversially the "queen" of the axiomatic sciences, there are different axiom sets in play. Our expectations for computational logico-mathematical modeling in the realm of human cognition don't exceed those in place for the application of such modeling to mathematics itself. A psychologist with such high expectations, to be consistent, would perhaps need to affirm the seemingly unpalatable view that the foundations of mathematics is less rigorous than the methods that ought to be in place in formal psychology.

6.1 Objection 2: But then what about organizations?

Despite (or in fact perhaps in light of) our rebuttal immediately above, a skeptic might object as follows: "If, as you indicate, new axioms, peculiar to the context and type of cognizers, are needed, your statements in the introduction that your methods for building the two models can be extended to modelling organizations, and in the conclusion that the two models substantiate this claim, is overly optimistic. Consider a parallel claim in another field, for instance the claim that a set of axioms couched in a logical framework capable of describing the physics of, say, the atom, with some additional assumptions, would be capable of modelling the physics of the molecule, including organic molecules (e.g., DNA), and that that in turn would be capable of modelling biological systems. Surely here a physicist would expect a — so to speak — 'deductive chain,' running from the atom on up."

This is a trenchant, humbling objection, certainly. But actually, thanks primarily to the recent groundbreaking work of mathematicians, logicians, and formally inclined physicists at the Rényi Institute of Mathematics, the axiomatization of physics is proceeding rather well (e.g., for special relativity, see Andréka, Madarász, Németi & Székely 2011). This axiomatization is based on separate sets of axioms for different parts of physics (special relativity, general relativity, quantum mechanics), and these axiom sets are not integrated in any way. But despite this "partitioned" state-of-affairs, the formal scientists in question are encouraged by the progress, heartened by the fact that different parts of the physical world *are* being formalized — albeit by different axiom sets. To be clear: We here say that the formal scientists in question are *heartened*; they are certainly not *satisfied* with separate axiom sets. That acknowledged, it seems quite reasonable to have expectations for the realm of cognizers that are not more demanding than those for the realm of physics; which is why we too are heartened. Our optimism about extending to organizations is based on the notion that as the union of axiom sets covering, say, relativity theory and quantum mechanics, is currently at least workable (though certainly not an ideal) situation, the parallel would hold for our approach and the modeling of organizations. Just as the size of objects in the case of physics is central, so apparently the type of agents in an organization would be key. And just as knowing what type of objects are to be modeled in the case of physics would be key, knowing what type of agents compose a given organization would be key. Ultimately, our hope is a counterpart to that which those in formal physics cling: The kinds of axiom sets we leverage in modeling and simulation will hopefully become part of a holistic theory of economic behavior.

Admittedly, it is reasonable for the reader to wonder, especially given empirical evidence supporting both models, under what circumstances one will see behavior consistent with BR1 vs. BR2. We concede that the mechanism by which agents select one logical model over the other is an aspect of an overall model we have not yet fully developed, and constitutes an area for future work. It would appear from empirical observation that this mechanism should incorporate prior knowledge, or the lack therof, on the part of *any* agent of the bi-pay auction structure; and, if at least one among the participants possesses significant prior knowledge, this mechanism should include to what degree an agent's actions are perceived by other participants to reveal this agent's knowledge and her consequent strategy. As our case in point, the forceful bidding by the agent perceived (by the other participants) as having knowledge, likely colored the beliefs of and decisions by the other participants in the \$20 experiment. This is unlike the \$100 experiment, where there didn't appear to be any "forceful" agents, and bidding proceded according to the expected trajectory above the prize threshold.

One might also seek details regarding how and why, in the BR1 model, agent reasoning switches from that based on the normal-auction logical model to the augmented one (which properly incorporates the subtle but profoundly different bi-pay auction). We have focused in this paper on detailing the cognitive models for before and after the "uh-oh" moment of realization. But from a computational standpoint, how does this revelation occur? This, too, is an area for future work.

6.2 Objection 3: What about the robustness of logico-computational models?

It is not unreasonable to imagine a researcher in the field of complex systems objecting to the nature of our models on the following grounds: "Models of complex systems tend to produce results that vary radically as a function of seemingly minor changes in assumptions about individual behavior. In light of this, individual-behavior assumptions should be kept extremely simple in an effort to preserve model robustness. Yet your techniques impose many complex assumptions about individual behavior. Your model risks falling victim to over-fitting and to vastly incorrect results that stem from your attempt to be realistic."

Although we acknowledge that the keep-it-simple principle should apply when building models of complex systems, we also believe that such concerns should be counter-balanced by realistic "high-fidelity" representation of individual behavior — when empirical evidence for such behavior exists. The approach presented herein allows for precisely that balance, and further, affords the capability to computationally deduce emergent behavior in complex systems given the formalized representation of individual behaviors and interactions. We know through a range of recent experiments (such as the bi-pay auction), about ways in which some individual humans deviate from fully rational behavior. We offer, here, a way to incorporate that new information when modeling human behavior and interaction. This approach may prove itself a worthy enhancement of the modeling of complex systems.

7 Conclusion and Future Work

Since the banking crisis of 2008 and the economic recession that followed, policy makers, pundits, and many in academia have decried the apparent fragility of the formal modeling upon which current economic theory is founded. By reducing economic behavior to formulas working in the aggregate, the complaint goes, we have abstracted human behavior away from the "cognitive reality" of real individuals, and as a result, these formulas fail. We have presented a computational modeling-and-simulation approach, which although formal, incorporates subtleties in real-world cognitive phenomena. Perhaps there was a time when computing capacity made this sort of approach infeasible in most practical contexts, and any appetite to build models that drilled down into (important) aspects of human cognition gave way to simplifying assumptions for pragmatic reasons. However, now, with increased computing power, we no longer have to confine the practice of formal modeling to the traditional approach that gives rise to the agent-less equations we commonly see in current economic theory. The methodology introduced herein, which is supported by the three Simonesque pillars of bounded rationality, simulation, and theorem proving, has the potential to explain in detail, and thus predict, economic phenomena, complete with justifications and verifications in the form of certified formal proofs.

As to future work, in addition to the items we discussed above that address specific issues of completeness of our model, we are in the process of taking two further steps forward. One, we are introducing probabilities and strength factors into our computational machinery This will not be difficult, since Slate was designed and engineered to enable such modeling (Bringsjord et al. 2008).[4] Our second step is to apply our methodology to situations in which many agents,

[4] And of course, probabilistic logics are a well-understood class of systems, at least when the underlying languages are extensional. For two examples, see (Halpern 1990) and (De Raedt & Kersting 2003).

each modeled at the detailed cognitive level described above, interact. We are seeking first to do this at the level of a type of organization with which we are familiar, as was Simon: a university.

References

Andréka, H., Madarász, J. X., Németi, I. & Székely, G. (2011), 'A Logic Road from Special Relativity to General Relativity', *Synthese* pp. 1–17.
URL: *http://dx.doi.org/10.1007/s11229-011-9914-8*

Arkoudas, K. & Bringsjord, S. (2007), 'Computers, Justification, and Mathematical Knowledge', *Minds and Machines* **17**(2), 185–202.
URL: *http://kryten.mm.rpi.edu/ka_sb_proofs_offprint.pdf*

Arkoudas, K. & Bringsjord, S. (2009), 'Propositional Attitudes and Causation', *International Journal of Software and Informatics* **3**(1), 47–65.
URL: *http://kryten.mm.rpi.edu/PRICAI_w_sequentcalc_041709.pdf*

Binmore, K., Rubinstein, A. & Wolinsky, A. (1986), 'The Nash Bargaining Solution in Economic Modelling', *The RAND Journal of Economics* **17**(2), 176–188.

Bringsjord, S. (2008), Declarative/Logic-Based Cognitive Modeling, *in* R. Sun, ed., 'The Handbook of Computational Psychology', Cambridge University Press, Cambridge, UK, pp. 127–169.
URL: *http://kryten.mm.rpi.edu/sb_lccm_ab-toc_031607.pdf*

Bringsjord, S. & Govindarajulu, N. S. (2011), 'In Defense of the Unprovability of the Church-Turing Thesis', *Journal of Unconventional Computing* **6**, 353–373. Preprint available at the url given here.
URL: *http://kryten.mm.rpi.edu/SB_NSG_CTTnotprovable_091510.pdf*

Bringsjord, S. & Govindarajulu, N. S. (2012), 'Given the Web, What is Intelligence, Really?', *Metaphilosophy* **43**(4), 464–479.
URL: *http://dx.doi.org/10.1111/j.1467-9973.2012.01760.x*

Bringsjord, S. & Govindarajulu, N. S. (2013), Toward a Modern Geography of Minds, Machines, and Math, *in* V. C. M¨ller, ed., 'Philosophy and Theory of Artificial Intelligence', Vol. 5 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*, Springer, New York, NY, pp. 151–165.
URL: *http://www.springerlink.com/content/hg712w4l23523xw5*

Bringsjord, S., Taylor, J., Shilliday, A., Clark, M. & Arkoudas, K. (2008), Slate: An Argument-Centered Intelligent Assistant to Human Reasoners, *in* F. Grasso, N. Green, R. Kibble & C. Reed, eds, 'Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8)', Patras, Greece, pp. 1–10.
URL: *http://kryten.mm.rpi.edu/Bringsjord_etal_Slate_cmna_crc_061708.pdf*

De Raedt, L. & Kersting, K. (2003), 'Probabilistic logic learning', *ACM SIGKDD Explorations Newsletter* **5**(1), 31–48.

Halpern, J. (1990), 'An analysis of first-order logics of probability', *Artificial Intelligence* **46**(3), 311–350.

Kolmogorov, A. & Uspenskii, V. (1958), 'On the Definition of an Algorithm', *Uspekhi Matematicheskikh Nauk* **13**(4), 3–28.

March, J. & Simon, H. (1958), *Organizations*, John Wiley and Sons, New York, NY.

Nash, J. (1953), 'Two-Person Cooperative Games', *Econometrica* **21**, 128–140.

NOVA (2010), Mind Over Money.
URL: *http://www.pbs.org/wgbh/nova/body/mind-over-money.html*

Russell, S. & Norvig, P. (2009), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ. Third edition.

Selten, R. (1999), What is Bounded Rationality?, *in* 'Dahlem Conference', Universitat Bonn, Adenauerallee 24-42, 53113 Bonn, Germany.

Simon, H. (1972), Theories of Bounded Rationality, *in* C. McGuire & R. Radner, eds, 'Decision and Organization', North-Holland, Amsterdam, The Netherlands, pp. 361–176.

Stickel, M. E. (2008), 'SNARK - SRI's New Automated Reasoning Kit'. Retrieved on July 26, 2013. http://www.ai.sri.com/ stickel/snark.html.
URL: *http://www.ai.sri.com/∼stickel/snark.html*

Fig. 9. Proof of $C8$**.** Using $[D5]$, $[D6]$, $[B2]$, $[B3]$, and $[I2]$, we infer that there is at least one bid above \$20.