SAMPLING AND UPDATING HIGHER ORDER BELIEFS IN DECISION-THEORETIC BARGAINING WITH FINITE INTERACTIVE EPISTEMOLOGIES

Paul Varkey and Piotr Gmytrasiewicz
Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan St., Chicago, U.S.A.

Keywords: Bilateral bargaining, Decision theory, Interactive epistemology, Higher order beliefs, Belief sampling.

Abstract: In this paper we study the sequential strategic interactive setting of bilateral, two-stage, seller-offers bargaining under uncertainty. We model the epistemology of the problem in a finite interactive decision-theoretic framework and solve it for three types of agents of successively increasing (epistemological) sophistication (i.e. capacity to represent and reason with higher orders of beliefs). We relax typical common knowledge assumptions, which, if made, would be sufficient to imply the existence of a, possibly unique, game-theoretic equilibrium solution. We observe and characterize a systematic monotonic relationship between an agent's beliefs and optimal behavior under a particular moment-based ordering of its beliefs. Based on this characterization, we present the *spread-accumulate* technique of sampling an agent's higher order belief by generating "evenly dispersed" beliefs for which we (pre)compute offline solutions. Higher order prior belief identification is then approximated to arbitrary precision by identifying a (previously solved) belief "closest" to the true belief. These methods immediately suggest a mechanism for achieving a balance between efficiency and the quality of the approximation – either by generating a large number of offline solutions or by allowing the agent to search online for a "closer" belief in the vicinity of best current solution.

1 INTRODUCTION

The central challenge that arises in the epistemological deliberations in strategic multi agent interactions under uncertainty, is representing and reasoning with the infinite interactive epistemology – first referred to in game-theoretic literature as "the infinite regress in reciprocal expectations" (Harsanyi, 1968).

Beginning with Harsanyi's seminal work on *Games under Incomplete Information* (Harsanyi, 1968), there is a long tradition of work and literature that has attempted to combine the game-theoretic notion of an equilibrium solution (Nash, 1950),(Nash, 1951) with a formal (probabilistic) calculus for representing and reasoning with the players' (individual, mutual or common) knowledge or, lack thereof. The solution concepts that have been proposed throughout the history of this research effort involve, (see first (Selten, 1975), then esp. (Fudenberg and Levine, 1981) and (Kreps and Wilson, 1982)), the proposal of a profile of strategies and contingent beliefs for each agent such that, given its beliefs, no agent has the incentive to deviate unilaterally. Further, the agents' beliefs about the objects of uncertainty (for e.g. the path of play, incl. esp. the previous choices of the other agents) have to be consistent with (i.e plausible) with respect to the joint profile of strategies.

No attempt will be made here to survey the field. It suffices to simply point out that usefulness (or, applicability) of game-theoretic solution concepts as a control paradigm in strategic multi agent interactions is unclear, regardless of their mathematical and abstract elegance and, oftentimes, powerful explanatory power. This is primarily due to the fact that there may be a multiplicity of equilibria for a particular game; although, there is a vast literature (see, for e.g., (Kreps, 1985), (Banks and Sobel, 1987), (Cho, 1987) and (Cho and Kreps, 1987)) discussing how one may refine the set of equilibria and select among them. Game-theoretic solution concepts have also been challenged on the basis of the fact that the common knowledge prerequisite (Aumann and Brandenburger, 1995) required to arrive at an equilibrium solutions may be unattainable.

Increasingly, (Bayesian) decision-theoretic and utility-theoretic approaches are becoming the dominant and normative paradigm for reasoning and decision making in settings under uncertainty. Recent

Varkey P. and Gmytrasiewicz P..
SAMPLING AND UPDATING HIGHER ORDER BELIEFS IN DECISION-THEORETIC BARGAINING WITH FINITE INTERACTIVE EPISTEMOLOGIES.
DOI: 10.5220/0003176901140123
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 114-123
ISBN: 978-989-8425-41-6

promising work extends this paradigm to interactive (i.e. multi agent) settings (Gmytrasiewicz and Doshi, 2005), (Doshi and Gmytrasiewicz, 2005). In these models, the recursive modeling of other agents' beliefs and reasoning is made explicit in a framework called the I-POMDP that extends classical POMDPs. Computable (i.e. finitely nested) instances of these models simply stop the recursive reflection after finite levels, allowing agents to make the best-possible decision with the information they chose to represent. In essence, they are (expected utility maximizing) decision-theoretic models that represent and reason with finite levels of the interactive belief hierarchy and make no (common knowledge) assumptions regarding the levels that are not modeled.

This finite interactive decision theoretic model is the assumed modeling framework throughout the current work, where we study a particular interactive sequential game – namely, two-stage seller-offers bargaining under incomplete information (Samuelson, 1984). It is known that this game has a unique Perfect Bayesian Equilibrium (Sobel and Takahashi, 1983), if it is asummed that the seller's belief about the buyer's valuation is commonly known.

In this work, we do not make the assumption that the seller's (first-order) belief is commonly known. Instead, we cast the problem in the finite interactive decision theoretic framework for which we derive optimal strategies for three types of agents of successively increasing epistemological sophistication.

Our first main contribution is the observation of a systematic regular (monotonic) relationship between the epistemology of the problem and the agents' optimal behavior. Secondly, this regularity is exploited to devise a belief generation scheme that generates beliefs that are "evenly dispersed" across an entire space of beliefs – equivalent to sampling the higher order belief in "evenly dispersed" locations. And, thirdly, solutions to these sample beliefs are precomputed offline for later use in the online stage – which consists of a binary search through the space of solved beliefs to identify the closest sample belief in order to more accurately approximately predict future behavior of the opponent.

In the next section, we describe the model(s) used throughout this paper and introduce necessary notations. In the following three sections, we describe, respectively, the deliberative reasoning process for each of the three strategy levels. Our main contributions with respect to (higher order) belief sampling, identification and updating are presented in the context of the discussion about the most sophisticated agent type studies in this paper – the L3-Buyer (Section 5). In the final section, we summarize our contributions, state ongoing work and discuss relevant open questions.

2 PRELIMINARIES

Throughout this paper, it is assumed that the seller's valuation $c = 0$ and that the buyer's valuation v is such that $0 \leq v \leq 1$. These are assumed to be commonly known; the exact value of v is the buyer's private information. The seller's belief about the buyer's valuation has the distribution $F(v) = v; 0 \leq v \leq 1$. We assume also that trade is feasible, i.e. that $c \leq v$. The mechanism is simple – the seller makes a first offer x_1 which the buyer may chose to accept; if the buyer rejects it, the seller makes a second and final offer x_2. The buyer strategic decision consists of choosing a decision boundary $d(x_1)$ – it accepts the first offer x_1 if $v \geq d(x_1)$.

If agreement is arrived at on the first day, the payoffs are x_1 and $v - x_1$ to the seller and buyer, respectively. If agreement is arrived at on the second day, the payoffs are $\delta \cdot x_2$ and $\delta \cdot (v - x_2)$, respectively. Else, the payoffs are 0 to either player. A discount factor, δ, is applied to the payoffs on the second day.

Agents may form other relevant beliefs and higher-order beliefs; for e.g. the seller may form a belief about the buyer's valuation, the buyer may form a second-order belief about the seller's first-order belief about its (i.e. the buyer's) valuation, etc. None of these beliefs are assumed to be commonly known.

2.1 Notations

The following notation will be used throughout.

$\mathbf{B_X(p)}$. Belief maintained by agent X (either seller S or buyer B) about p, where p is the object of the agent's belief and may be a ground proposition or another agent's belief about something (this will be clear from the context).

$\mathbf{U(s)}$. A uniform belief supported over a space s, where s may be a (finite or countable) set of ground propositions or a set of beliefs.

$\mathbf{E[v]}$. Expected value of random variable v.

$\mathbf{p_{Accept}(x)}$. The probability that offer x is accepted.

$\mathbf{p_{Reject}(x)}$. The probability that offer x is rejected (equal to $1 - p_{Accept}(x)$).

$\pi_1(\cdot)$. The expected utility function for the entire (2-stage) sequential bargaining game. We denote $\arg\max \pi_1(\cdot)$ by $\Pi_1(\cdot)$.

$\pi_2(\cdot)$. The expected utility function for the last (i.e. second) stage of the bargaining game. We denote

$\arg\max \pi_2(\cdot)$ by $\Pi_2(\cdot)$. The influence of an agent's belief on its objective function is indicated by appending the belief as a superscript to the expected utility function as well as to the probabilities of an offer being accepted or rejected.

2.2 The Epistemic Setup

Using the notations introduced above, we now define the three types of successively more sophisticated agents that are studied in this paper along with the interactive epistemology that arises due to interactions between these agents. Please see Figure 1 for a graphical illustration of the same.

L1-Buyer believes that the seller's first offer x_1 is uniformly selected from $(0,1)$ and that its second offer is uniformly distributed between 0 (the lowest possible) and the first offer, i.e. $B_B(x_2) \sim U(0,x_1)$; in other words, that the seller's second offer is some arbitrary amount lesser than the first.

L2-Seller believes that the buyer's belief about x_2 given x_1 is that it is supported on $(0,x_1)$ – although, the seller does not "know" that this belief is uniformly distributed. Therefore, it maintains a higher (second) order belief that is supported on some subset of all possible beliefs about x_2 given x_1 that are supported on $(0,x_1)$, i.e. $B_S(B_B(x_2)) \sim U(\Delta(0,x_1))$. The seller can express this higher-order belief over a space of beliefs as a lower order belief about the buyer's expectation. Here we say, for example, that the seller's belief about the buyer's expectation of x_2 given x_1 – $B_S(E_B[x_2|x_1])$ – is uniformly distributed on $(0,x_1)$. Note that the L2-Seller's belief about the buyer's type is characterized here by its (second order) belief about the buyer's first order belief about the second offer, or, equivalently, by its belief about the buyer's conditional expectation.

L3-Buyer The buyer believes that the seller's belief about $E_B[x_2|x_1]$ is supported on $(0,x_1)$ – although, the buyer does not "know" that this belief is uniformly distributed; in other words, the buyer does not "know" that the seller is uniformly uncertain about the buyer's belief about x_2 given x_1. Therefore, this buyer maintains a higher order belief that is supported on some subset of all possible beliefs that the seller may maintain about $E_B[x_2|x_1]$.

3 L1-BUYER

The L1-Buyer accepts the first offer x_1 if the immediate profit is not lesser than the expected profit from

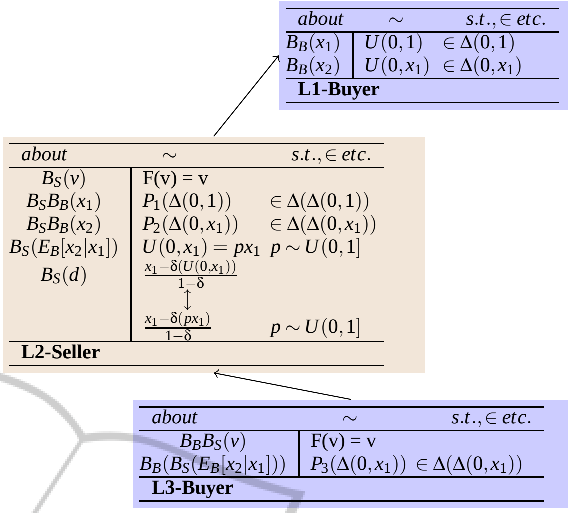


Figure 1: An interactive epistemology for bilateral bargaining.

the next turn, i.e. if

$$v - x_1 \geq \delta(v - E[x_2|x_1])$$
$$\Rightarrow v - x_1 \geq \delta(v - x_1/2)$$

i.e. if

$$v \geq \frac{x_1 - \delta\frac{1}{2}x_1}{1-\delta} =: d, \text{the decision cutoff}$$

4 L2-SELLER

Note that the L1-Buyer believes that the second offer x_2 is selected from some distribution in $\Delta(0,x_1)$ – namely, the space of all distributions supported on $(0,x_1)$. Therefore, given x_1, it can calculate the expected value of the second offer $E_B[x_2|x_1]$ – which is all it needs to compute the optimal decison boundary. For e.g., if the L1-Buyer believes that $x_2 \sim U(0,x_1)$ (as is actually the case as shown in the previous section), then $E_B[x_2|x_1] = \frac{1}{2}\cdot x_1$. In general, $E_B[x_2|x_1]$ has the form $p \cdot x_1$ (for $0 \leq p \leq 1$). Therefore, from the L2-Seller's perspective, only $B_S(E_B[x_2|x_1])$ – i.e. its beliefs about $E_B[x_2|x_1]$ – matter when it reasons about the buyer's reasoning process following the rejection of the first offer. In the epistemics considered here (see Figure 1), the seller believes that $E_B[x_2|x_1] \sim U(0,x_1)$ (or, $E_B[x_2|x_1] = p \cdot x_1$ where $p \sim (0,1]$). The seller's best-response consists of the computation of x_1^* such that

$$x_1^* \in \arg\max_{x_1} \pi_1^{B_S(E_B[x_2|x_1])}(x_1)$$
$$= \arg\max_{x_1} \pi_1^{U(0,x_1)}(x_1)$$

$$= \arg\max_{x_1} \Big(p^{U(0,x_1)}_{Accept}(x_1)\cdot x_1 + \delta\cdot p^{U(0,x_1)}_{Reject}(x_1)\cdot \Pi_2^{U(0,x_1)}(x_1)\Big)$$

The influence of the seller's belief about $E_B[x_2|x_1]$ on the values of the obective function at each stage and on the probabilities of acceptance is indicated by including the belief as the superscript of such values.

4.1 Uniform Belief with Finite Support

We now focus our attention on a specific epistemic setup to fix intuition. Assume that $B_S(E_B[x_2|x_1])$ has a distribution U_3 s.t.:

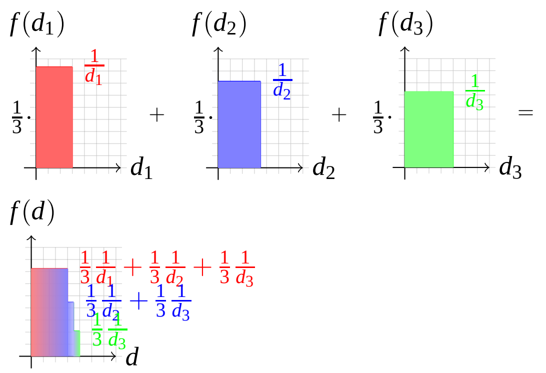
$$B_S(E_B[x_2|x_1]) \sim U_3 \sim U\left(\frac{x_1}{4}, \frac{x_1}{2}, \frac{3x_1}{4}\right)$$

The purpose of this setup is two-fold – first, it may be seen as a simplified approximate version of the seller's actual belief and, second, it may also be thought of as a means of gaining useful insight into the epistemic deliberative and computational challenges inherent in the original problem.

Denote the corresponding values of the buyer's decision boundary d as d_1, d_2 and d_3, where

$$d_1 = \frac{x_1 - \delta\frac{1}{4}x_1}{1-\delta},\ d_2 = \frac{x_1 - \delta\frac{1}{2}x_1}{1-\delta}\ \&\ d_3 = \frac{x_1 - \delta\frac{3}{4}x_1}{1-\delta}$$

If a particular offer x_1 is rejected, the resultant posterior density from the seller's belief update is a weighted sum of three densities, represented pictorially as:



The probability that x_1 will be accepted can be expressed as:

$$p^{U_3}_{Accept}(x_1) = \frac{1}{3}\left[\frac{F(1)-F(d_1)}{F(1)}\right] + \frac{1}{3}\left[\frac{F(1)-F(d_2)}{F(1)}\right] + \frac{1}{3}\left[\frac{F(1)-F(d_3)}{F(1)}\right]$$
$$= \frac{1}{3}(1-d_1) + \frac{1}{3}(1-d_2) + \frac{1}{3}(1-d_3)$$

Now, the the seller assigns the following probability to the buyer accepting the second offer:

$$p^{U_3}_{Accept}(x_2) = \begin{cases} \frac{1}{3}\frac{F(d_1)-F(x_2)}{F(d_1)} + \frac{1}{3}\frac{F(d_2)-F(x_2)}{F(d_2)} + \frac{1}{3}\frac{F(d_3)-F(x_2)}{F(d_3)} & \text{if } x_2 \le d_1 \\ \frac{1}{3}\frac{F(d_2)-F(x_2)}{F(d_2)} + \frac{1}{3}\frac{F(d_3)-F(x_2)}{F(d_3)} & \text{if } d_1 \le x_2 \le d_2 \\ \frac{1}{3}\frac{F(d_3)-F(x_2)}{F(d_3)} & \text{if } d_2 \le x_2 \le d_3 \end{cases}$$

Accordingly, the seller's second stage objective function (i.e. expected profit from offering some x_2 after x_1 has been rejected) becomes:

$$\pi_2(x_2) = \begin{cases} \frac{1}{3}\frac{F(d_1)-F(x_2)}{F(d_1)}x_2 + \frac{1}{3}\frac{F(d_2)-F(x_2)}{F(d_2)}x_2 + \frac{1}{3}\frac{F(d_3)-F(x_2)}{F(d_3)}x_2 & \text{if } x_2 \le d_1 \\ \frac{1}{3}\frac{F(d_2)-F(x_2)}{F(d_2)}x_2 + \frac{1}{3}\frac{F(d_3)-F(x_2)}{F(d_3)}x_2 & \text{if } d_1 \le x_2 \le d_2 \\ \frac{1}{3}\frac{F(d_3)-F(x_2)}{F(d_3)}x_2 & \text{if } d_2 \le x_2 \le d_3 \end{cases}$$

Let $$X_2(x_1) = \arg\max_{x_2} \pi_2(x_2)$$

and $$\Pi_2(x_1) = \max_{x_2} \pi_2(x_2)$$

respectively.

We first obtain both of these as functions of x_1 from the piecewise first-order conditions for $\pi_2(x_2)$. Then, we express the seller's main (first stage) objective function as:

$$\pi_1(x_1) = \left(\frac{1}{3}(1-d_1) + \frac{1}{3}(1-d_2) + \frac{1}{3}(1-d_3)\right)\cdot x_1 + \delta\cdot\left(\frac{1}{3}d_1 + \frac{1}{3}d_2 + \frac{1}{3}d_3\right)\cdot \Pi_2(x_1)$$

The first-order conditions for $\pi_1(x_1)$ provide us the seller's optimal first offer X_1 as 0.3831693366 (from which we can easily compute the the optimal second offer X_2 and the expected optimal profit Π_1).

The algorithm for the general case (for evenly distributed uniform discretized beliefs of finite support) is provided by Procedure X1X2Pi1.

Procedure: X1X2Pi1(N,δ,p).

Input: N ← the number of samples, δ ← the discount factor and p ← an array (of size N) of probabilities

Output: $X_1 \leftarrow \texttt{argmax}_{x_1}\pi_1(x_1)$,
$X_2(X_1) \leftarrow X_2(x_1 = X_1)$ and
$\Pi_1(X_1) \leftarrow \pi_1(x_1 = X_1)$

// Given the number of samples, the discount factor and a prior belief sampling, returns the seller's optimal offer schedule and expected optimal profit for the case where $B_S(v) := F(v) = v$

begin
  // D ← decision cutoffs
  $D[0] \leftarrow 0$
  for *i from* 1 *to* $N-1$ **do**
    $D[i] \leftarrow \frac{x_1 - \delta(\frac{N-i}{N})x_1}{1-\delta}$
    // $C\pi_2 \leftarrow$ piecewise constrain intervals
    $C\pi_2[i] \leftarrow D[i-1] < x_2$ and $x_2 \le D[i]$
    // $\pi_2 \leftarrow$ final stage piecewise objective function
    $$\pi_2[i] \leftarrow \sum_{j=i}^{N-1} p[N-j]\frac{D[j]-x_2}{D[j]}x_2$$
    // $\frac{\partial\pi_2}{\partial x_2} \leftarrow$ piecewise partial derivative of π_2 w.r.t x_2
    $\frac{\partial\pi_2}{\partial x_2}[i] = \frac{\partial\pi_2[i]}{\partial x_2}$
  end
  // Initializations
  $X_2 \leftarrow 0$; $\Pi_2 \leftarrow 0$
  // Compute $X_2(x_1) \leftarrow \texttt{argmax}_{x_2}\pi_2(x_2)$ and $\Pi_2(x_1) \leftarrow \texttt{max}_{x_2}\pi_2(x_2)$
  for *i from* 1 *to* $N-1$ **do**
    $x_2[i] \leftarrow$
    $solve(\frac{\partial\pi_2}{\partial x_2}[i], x_2)$ assuming $C\pi_2[i]$
    Let $x_1 > 0$ and $x_2 \leftarrow x_2[i]$
    if $C\pi_2[i]$ *and* $\pi_2[i] > \Pi_2$ **then**
      $X_2 \leftarrow x_2$
      $\Pi_2 \leftarrow \pi_2[i](X_2)$
    end
  end
  // $\pi_1 \leftarrow$ the main (first stage) objective function
  $$\pi_1(x_1) \leftarrow \left(\sum_{j=1}^{N-1} p[N-j](1-D[j])\right)x_1 + \delta\left(\sum_{j=1}^{N-1} p[N-j]D[j]\right)\Pi_2$$
  $X_1 \leftarrow solve(\frac{\partial\pi_1(x_1)}{\partial x_1}, x_1)$
  return X_1, $X_2(x_1 = X_1)$, $\pi_1(x_1 = X_1)$
end

We now highlight a few insights gained:

1. The optimal offers and optimal expected profits for the cases when each of the sample points considered here is the certain case (i.e. w.p. 1), is shown in Table 1.

Table 1: Optimal offers and expected profits when seller "knows" $E_B[x_2|x_1]$.

$\beta^\dagger$	$x_1^{\beta,*} = \arg\max_{x_1}\Pi_1^{\beta}(x_1)$	$\Pi_1^{\beta}(x_1^{\beta,*})$
$\frac{x_1}{4}$	0.3453858608	0.1726929304
$\frac{x_1}{2}$	0.3874092010	0.1937046005
$\frac{3x_1}{4}$	0.4581245526	0.2290622763

$^\dagger\beta$ denotes $B_S(E_B[x_2|x_1])$

From this table, we notice, in particular, that $x_1^{\frac{x_1}{2},*} \neq x_1^{U_3,*}$ (though $E[U_3] = \frac{x_1}{2}$), i.e. that

$$x_1^{\frac{x_1}{2},*} = 0.3874092010 \neq x_1^{U_3,*} = 0.3831693366$$

though they are "very close". This indicates at least that, though $E[U(0,x_1)] = \frac{x_1}{2}$,*it is not necessarily the case that*

$$x_1^{U(0,x_1),*} = x_1^{\frac{x_1}{2},*}$$

In other words, *the solution to an optimization problem parametrized by a random variable may, at best, only be approximated by the solution to the related optimization problem parametrized by the expected value of the parameter.*

Table 2: Optimal offers and expected profits for uniform discrete seller beliefs.

N	$x_1^{U_N,*} = \arg\max_{x_1}\Pi_1^{U_N}(x_1)$	$\Pi_1^{U_N}(x_1^{U_N,*})$
5	0.3822887563	0.1911443782
10	0.3805058597	0.1902529299
20	0.3796076053	0.1898038026
50	0.3790678491	0.1895339246
100	0.3788879236	0.1894439618
200	0.3787979719	0.1893989860
400	0.3787529981	0.1893764991

2. The solutions for evenly distributed uniform discretized beliefs of varying supports is collected in Table 2. As we increase the support of the seller's belief (i.e. as N grows), we observe a Cauchy-like behavior in the corresponding optimal (first) offers. In ongoing work, we are attempting to establish whether these values do indeed converge (as they seem to) and whether they converge to $x_1^{U(0,x_1),*}$ (as hoped for).

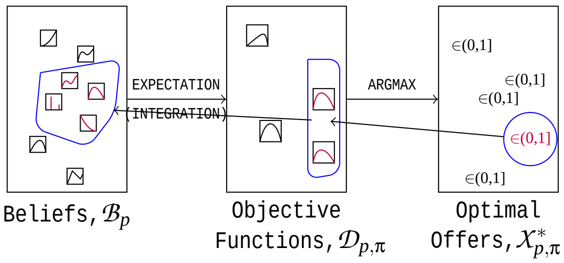


Figure 2: The mapping from beliefs to objective functions to optimal actions.

5 L3-BUYER

The L3-type buyer's best response, i.e. either to reject or accept the first offer x_1, consists of computing the optimal decision boundary d^*.

From the L3-Buyer's perspective, its belief about the seller's belief about $E_B[x_2|x_1]$ (or, equivalently its belief about the seller's belief about p) – namely, $B_B B_S(E_B[x_2|x_1])$ (or, $B_B B_S(p)$) – constitutes the crucial epistemic component that influences the computation of the optimal decision boundary.

In the simplest case, the L3-Buyer "knows" $B_S(p)$ with certainty. In this case, it can solve the seller's optimization problem (for e.g., just like the seller does, through sampling) to form expectations about x_2 which, in turn, it uses to compute d^*.

In the general case, the buyer needs a method whereby it can use the seller's first offer as an *informative signal* to update $B_B B_S(p)$ – i.e. to compute $B_B B_S(p|x_1)$. The buyer deems a particular seller belief implausible if a seller with such a belief would have never sent the recieved (i.e. actual) first offer. This constitutes the central contribution of the current work and is discussed in detail next.

5.1 Higher Order Belief Identification and Refinement (Update)

Figure 2 represents the mapping from the seller's belief to the objective function to the optimal (first) offer. The L3-Buyer is interested in the inverses of these mappings in order to refine the support of its belief about the seller's belief.

The L2-Seller's objective function is parametrized by its belief about $E_B[x_2|x_1]$ – which it expresses as a belief about p (recall that $E_B[x_2|x_1] = p \cdot x_1$). It obtains the appropriate form of the objective function by computing its expected value, i.e. by *integrating over its belief about p* (see Figure 2).

Let $\mathcal{B}_p$ be the class of all possible beliefs about p. The seller's objective function π is characterized by parameter p (and, henceforth, will be written as π_p). Let $\mathcal{D}_{p,\pi_p}$ be the class of all objective functions generated by integrating π over p for all possible beliefs about p (i.e., over all elements of $\mathcal{B}_p$). The seller obtains the optimal (first) offer from the first-order conditions for the objective function.

In general, it is true that there are an uncountable number of functions that share the same *argmax* in a given interval. But, for the buyer, the relevant question is

Question 1. Is the mapping from the seller's *possible* objective functions, $\mathcal{D}_{p,\pi}$, to the optimal offers, $\mathcal{X}^*_{p,\pi_p}$, injective?

A related and important question is

Question 2. Is the mapping from the seller's beliefs to objective functions injective?

Note that the buyer's update process consists of contingent-factual reasoning which may involve inverting the two discussed mappings (Figure 2). Such inversions necessitate that the maps be injective.

Unfortunately, these maps are *not injective*; an observation which can be established by a few examples. So, then, how does the buyer update its belief?

It is in the context of this problem that we present the main contribution of this paper. Our approach involves looking at the buyer's belief update problem in the other (i.e. the forward) direction – i.e. from the seller' beliefs to its actions. Establishing that the optimal offers (schedules) are monotonic w.r.t. some suitable ordering of the seller's beliefs would point to a means of identifying the seller's next offer (with arbitrary precision) by searching for a *close-enough* belief that would explain the seller's first offer. We now proceed to expound this idea in depth.

5.2 (μ,σ)-Ordered D.F.s

Consider the seller's objective function $\pi_p : [0,1] \rightarrow \mathcal{R}^+$. π_p is continuous and differentiable in $[0,1]$ and attains its maximum value at the only saddle-point in the same interval. π_p is parametrized by a random variable p (where $0 \leq p \leq 1$). If the distribution F of p is known, one can calculate the expectation of the function π_p :

$$E_F(\pi_p) = \int_0^1 \pi_p dF$$

Since the buyer does not, in general, know F, we assume that the distribution function of p comes from a family of distribution functions $\mathcal{F}_p$. Here, we define a particular ordering for classes of distributions.

Order$_{(\mu,\sigma)}(\mathcal{F})$. Given a class of distribution functions $\mathcal{F}$, *Order*$_{(\mu,\sigma)}(\mathcal{F})$ is a (total) ordering of $\mathcal{F}$ based on the expectation (i.e. mean value) of the member distributions, where, if possible, ties are broken by further (totally) ordering based on the variance of the member distributions. In symbols,

$$
\begin{aligned}
F &\prec_{Order_{(\mu,\sigma)}(\mathcal{F})} G \\
&\text{if } \mu(F) < \mu(G) \text{ or} \\
&\quad \mu(F) = \mu(G) \text{ and } \sigma(F) > \sigma(G) \\
&\forall \; F, G \in \mathcal{F}
\end{aligned}
$$

Now consider a sequence of distribution functions $F_1, F_2, F_3, \ldots$ from $\mathcal{F}_p$ that respects $Order_{(\mu,\sigma)}(\mathcal{F}_p)$ in addition to satisfying an additional technical requirement that the F_n are "not too close".

We obtain empirical evidence that indicates that, under such an ordering, the distribution functions F_n have a monotonic influence on the saddle-point of $E_{F_n}(\pi_p)$, i.e.:

$$\texttt{argmax}\ E_{F_n}(\pi_p) = \int_0^1 \pi_p dF_n$$

$$\leq$$

$$\texttt{argmax}\ E_{F_m}(\pi_p) = \int_0^1 \pi_p dF_m$$

whenever $1 \leq n < m$.

This seems to be the case because of the fact that p exerts a monotonic influence on $\texttt{argmax}\ \pi_p$ (see Table 1). Clearly, this observation is specific to the problem and the mechanism under consideration and it is not immediate whether it is true in general. The statement and proof of a general result of this type is a subject of current work.

In any case, the empirically established monotonic behavior observed in the current settings provides sufficient grounds for the development of the higher order belief identification method that is outlined in the next subsection.

5.3 (μ,σ)-ordered *Spread-Accumulate* Sample Generation Method for Higher Order Belief Sampling

Based on the observations of the previous section, the buyer can pre-compute the seller's optimal first offer for various seller belief settings[1] that are "appropriately evenly dispersed" according to the (μ,σ)-ordering. Then, when it recieves a particular first offer from the seller, it can first find the nearest pre-computed belief setting that explains this offer and, if necessary, search in the (μ,σ)-vicinity of this belief for a "better explanation".

First, we need a systematic method of sampling to choose and solve for a set of beliefs that are "appropriately evenly dispersed" across the entire space of possible beliefs (in the buyer's support). At the least, the method has to select a set of distributions with mean values that range from one extreme end of the support interval to the other. Further, for a particular mean, it has to select distributions with a range of variances – from low to medium to high.

We have devised exactly such a method – and we call it the *spread-accumulate* sample generation method. It consists of the following steps:

1. Discretize the support of the seller's belief (about p) uniformly into N intervals. Here $p \in (0,1]$; therefore, the discretized beliefs will be supported on $\left\{\frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}\right\}$.
2. Generate discrete distributions each of which has a mean exactly equal to one of the support points (for all values from $\left\{\frac{1}{N}\right\}$ to $\left\{\frac{N-1}{N}\right\}$), for different settings of the variance (from low to high). Two distinct kinds of cases can be identified: **Extreme points (2 cases).** The entire mass of the distribution ($= 1$) is concentrated on one of the extreme points of the supporing set. The variance setting cannot be changed (i.e. decreased) any further without shifting the mean. **Interior points ($(\mathbf{N-3})(\mathbf{N-1})$ cases).** For a particular mean, we first choose and solve for the discrete belief which concentrates all its mass on the mean value. Then, this being the key step of the *spread-accumulate* method, we select distributions by alternatively successively spreading out the probability mass from the previous distributions to neighbouring points and then accumulating some probability mass from the interior points to the most extreme points with positive mass. An example will illuminate: Let $N = 7$. Let the desired mean be $\frac{4}{7}$. Then, the *spread-accumulate* method generates discrete distributions on the support $\left\{\frac{1}{7}, \frac{2}{7}, \ldots, \frac{6}{7}\right\}$ in the following order (only probability masses are specified for the points in the support in the natural order):

$$
\begin{array}{cccccc}
0, & 0, & 0, & 1, & 0, & 0 \\
0, & 0, & \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3}, & 0 \\
0, & 0, & \frac{1}{2}, & 0 & \frac{1}{2}, & 0 \\
0, & \frac{1}{4}, & \frac{1}{4}, & 0, & \frac{1}{4}, & \frac{1}{4} \\
0, & \frac{1}{2}, & 0, & 0, & 0, & \frac{1}{2} \\
\frac{2}{5}, & 0, & 0, & 0, & 0, & \frac{3}{5}
\end{array}
$$

The buyer solves the seller's optimization problem for each of the generated sample beliefs and

[1] For convenience, we only consider discrete beliefs

stores the corresponding optimal schedules (the first and second offers) in 2D arrays. The whole process – *spread-accumulate* sample generation and optimal schedule computation – is described in Procedure L3BuyerOfflineX1X2Pi1. Row i corresponds to beliefs with mean$=\frac{i}{N}$. The column entries correspond to different settings of the variance – from low (entry 1) to high (entry $n-1$).

5.4 Online Computation of d^*

The offline pre-computation of the seller's optimal schedule for various seller belief settings, described in the previous section, enables the buyer to identify the "closest" belief in the (μ,σ)-ordering to the seller's actual belief (this is exactly the belief that is associated with the optimal first offer closest to the seller's actual first offer). In the offline step, the buyer stores results into the 2D array in the naturally available monotonic ordering – this is exploited here to implement the buyer's closest-belief-identification procedure as a *binary search*. The buyer can then also, optionally, perform a refined search (through further sampling and computation) in this vicinity to identify a "closer" belief. Once the buyer finds a "close enough" solution (to arbitrary precision), it uses the associated optimal second offer to compute d^*. This process is formally outlined in Procedure L3BuyerGetX1RetD*.

5.5 An Example

We consider the case of an L2Seller, Sam bargaining with an L3Buyer, Bob. (We also assume that $\delta = 0.6$ throughout this example). Sam models the buyer as an L1-type buyer (cf. Section) and has the following belief about the buyer's expectation of the second offer given the first: $B_S(E_B[x_2|x_1]) = p \cdot x_1$ where p is supported on $[2/5, 4/5]$ with respective probabilites $[2/3, 1/3]$. This may be interpreted as Sam's belief that Bob is twice as likely to expect a huge decrease in the second offer, corresponding to $P_{B_S(E_B[x_2|x_1])}(p = 2/5) = 2/3$, than not. Sam's problem may be solved to obtain the optimal offers as:

$$x1^*_{Sam} = 0.39 \text{ and } x2^*_{Sam} = 0.32$$

Now, we consider Bob's offline step. Say that Bob chooses discretization paramater, N, to be 6. Bob's solution for all the 17 $(= 2 + (N-3)(N-1))$ sampled belief points are recorded in Table 3 where each entry (i, j) is a 2-tuple $(X1[i, j], X2[i, j])$, such that $X1[i, j]$ and $X2[i, j]$ comprise the solution corresponding to the belief sample with mean $\frac{i}{N}$ and variance corresponding to the jth smallest for that particular mean value (generated according to *spread-accumulate* sampling).

Procedure: L3BuyerOfflineX1X2Pi1(N,δ).

Input: N ← the number of discretization intervals
    and δ ← the discount factor
Output: X_1, X_2 and $\Pi_1 \leftarrow (N-1) \times (N-1)$ 2D arrays representing, respectively, the optimal schedule and optimal expected profit to the seller for different settings of seller beliefs
// Procedure X1X2Pi1 will be invoked throughout
begin
  // a ← a size $(N-1)$ probability vector
  $a \leftarrow \{1, 0, ..., 0\}$
  $(X_1[1,1], X_2[1,1], Pi_1[1,1]) = X1X2Pi1(N, \delta, a)$
  for *m from* 2 *to* $N-2$ **do**
    $a \leftarrow \{0, 0, ..., 0\}$, $a[m] = 1$
    $(X_1[m,1], X_2[m,1], Pi_1[m,1]) = X1X2Pi1(N, \delta, a)$
    $a \leftarrow \{0, 0, ..., 0\}$,
    $a[m-1] = a[m] = a[m+1] = \frac{1}{3}$
    $(X_1[m,2], X_2[m,2], Pi_1[m,2]) = X1X2Pi1(N, \delta, a)$
    $a \leftarrow \{0, 0, ..., 0\}$, $a[m-1] = a[m+1] = \frac{1}{2}$
    $(X_1[m,3], X_2[m,3], Pi_1[m,3]) = X1X2Pi1(N, \delta, a)$
    $c = \min(m-2, N-2-m)$
    for *n from* 1 *to c* **do**
      $a \leftarrow \{0, 0, ..., 0\}$, $a[m+n+1] = a[m+n] = a[m-n] = a[m-n-1] = \frac{1}{4}$
      $(X_1[m, 2n+2], X_2[m, 2n+2], Pi_1[m, 2n+2]) = X1X2Pi1(N, \delta, a)$
      $a \leftarrow \{0, 0, ..., 0\}$,
      $a[m+n+1] = a[m-n-1] = \frac{1}{2}$
      $(X_1[m, 2n+3], X_2[m, 2n+3], Pi_1[m, 2n+3]) = X1X2Pi1(N, \delta, a)$
    end
    if $m < \frac{N}{2}$ **then**
      for *l from* 2m *to* $N-1$ **do**
        $a \leftarrow \{0, 0, ..., 0\}$, $a[1] = \frac{m-l}{1-l}$ and $a[l] = 1 - a[1]$
        $(X_1[m,l], X_2[m,l], Pi_1[m,l]) = X1X2Pi1(N, \delta, a)$
      end
    else
      for *l from* 1 *to* $2m-N$ **do**
        $a \leftarrow \{0, 0, ..., 0\}$,
        $a[2m-N-l+1] = \frac{m-N+l}{2m-2N-l+2}$ and
        $a[N-1] = 1 - a[2m-N-l+1]$
        $(X_1[m, 2N-2m+l-1], X_2[m, 2N-2m+l-1], Pi_1[m, 2N-2m+l-1]) = X1X2Pi1(N, \delta, a)$
      end
    end
  end
  $a \leftarrow \{0, 0, ..., 1\}$
  $(X_1[N-1,1], X_2[N-1,1], Pi_1[N-1,1]) = X1X2Pi1(N, \delta, a)$
  return X_1, X_2, Π_1
end

Table 3: Output of Procedure L3BuyerOfflineX1X2Pi1 for Bob.

$i\downarrow$, $j\rightarrow$	1	2	3	4	5
1	0.3354, 0.3773	–	–	–	–
2	0.3571, 0.3571	0.3555, 0.3518	0.3547, 0.3492	0.3517, 0.3391	0.3478, 0.3260
3	0.3874, 0.3389	0.3855, 0.3327	0.3846, 0.3296	0.3803, 0.3155	0.3764, 0.3025
4	0.4301, 0.3225	0.4277, 0.3148	0.4266, 0.3111	0.4242, 0.3030	0.4224, 0.2970
5	0.4923, 0.3076	–	–	–	–

Procedure: L3BuyerGetX1RetD*(x1,N,X1,X2,δ).

Input: x1 ← the (L2-type) seller's first offer, N ← the dimension of the precomputed 2D arrays, $X1$ and $X2$ and δ ← the discount factor
Output: s* ← the (L3-type) buyer's decision boundary

```
// The set of beliefs with the closest
   mean value if first identified using
   binary search; then, the belief with
   the closest variance is identified
   using linear search through this class
```

begin
- $l \leftarrow 1$
- $u \leftarrow N-1$
- **for** *i from 1 to ceil*$(ln(N)+1)$ **do**
  - **if** $x1 \geq X1\left[ceil\left(\frac{l+u}{2}\right),1\right]$ **then**
    - $l \leftarrow floor\left(\frac{l+u}{2}\right)$
  - **else**
    - $u \leftarrow ceil\left(\frac{l+u}{2}\right)$
  - **end**
- **end**
- $i_1 \leftarrow l$
- $i_2 \leftarrow 1$
- $\varepsilon \leftarrow |x2 - X1[i_1, i_2]|$
- **for** *i from 1 to* $N-1$ **do**
  - **if** $|X1[u,i] - x_1| \leq \varepsilon$ **then**
    - $i_1 \leftarrow u$
    - $i_2 \leftarrow i$
    - $\varepsilon \leftarrow |X1[u,i] - x_1|$
  - **end**
- **end**
- $d^* \leftarrow \frac{x1-\delta X2[i_1,i_2]}{1-\delta}$
- `return` d^*

end

Procedure L3BuyerGetX1RetD*'s trace with respect to Table 3 shows that, after three binary search steps and five linear search steps, the "closest" precomputed first offer given the actual value of 0.39 is 0.3874, corresponding to $X1[3,1]$. The corresponding best estimate of the second offer, $X2[3,1]$, is 0.3389, which is very close to the actual second offer, 0.32. The cutoff that is computed is

$$\frac{x1 - \delta \cdot X2[3,1]}{1-\delta} = \frac{0.39 - 0.6 \cdot 0.3389}{1 - 0.6} = 0.46665$$

while the optimal cutoff is

$$\frac{x1 - \delta \cdot x2}{1-\delta} = \frac{0.39 - 0.6 \cdot 0.32}{1 - 0.6} = 0.495$$

Notice, importantly, that only a *significantly small fraction of buyers*, those with valuations in the range $[0.46665, 0.495]$, make the *wrong* decision to accept the first offer. In particular, for e.g., in the assumed commonly known epistemology analysed here, the bueyrs are uniformly distributed in $[0,1]$, implying that 97.165% *of the buyers make the right decision*.

6 CONTRIBUTIONS AND ONGOING WORK

In the course of analyzing the L2-Seller (cf. Section 4), we graphically illustrated the step-function shape of a (Bayesian) updated belief density for an agent that maintains a discrete uniform prior belief and uses the opponent's signal as a screening device. In ongoing work, we are working on a *generalized belief update* method for computing posterior densities using a screening signal for general (i.e. non-uniform) discrete prior beliefs. In addition, we are investigating whether the optimal strategies of the L2-Seller converge when we increase the number of samples used to represent its discrete uniform prior belief.

In Section 5, we analyzed the L3-Buyer and presented the central contributions of this paper. We made an important observation about the regular (monotonic) influence of a particular ordering, namely, the (μ,σ)**-ordering**, of distribution functions of a random variable on the (maximal) saddle-point of an objective function that is paramatrized by that variable. In this context, we presented a question that will be the subject of future work: *Does the observed regular (monotonic) influence occur because the random variable itself exerts a similar monotonic influence?* A second question that is a foundational to extending these results is, *In general, when do the (central) moments suffice in completely characterizing the influence of the epistemology of the problem on optimal behavior? And, when they do, how many (central) moments suffice?*

Next, we motivated and developed the ***spread-accumulate*** (belief) sample generation method. This method enables the generation of "evenly dispersed" samples from the higher order belief space. These samples are then pre-solved offline to facilitate online binary-search for the "closest" belief instead of doing exact belief update (which may be analyticaly unachievable and numerically intractable). In this manner, we are able to realize (approximate) higher-order belief update. The monotonicity property of the (μ, σ)-ordering is exploited to implement the binary search through the space of (sampled and pre-solved) beliefs to identify the closest representative.

We then proposed that this higher-order belief update (or, approximate identification) scheme may be seen as an online-refinement based realization of *bounded rationality*. If the precomputed solutions are not "close enough" (within some desired precision), the agent can perform further online computations and fine-grained search within the (μ, σ)-vicinity of the current best solution. The agent can increase the quality of its *offline* solutions during its down-time as well as seek a better approximations *online*.

6.1 Conclusions

In conclusion, we have demonstrated the usefulness and epistemological modeling power of the *framework of multi-agent decision-theoretic reasoning and sequential planning with finite interactive epistemologies* (Gmytrasiewicz and Doshi, 2005) for the real-world problem of bilateral bargaining. We focussed on the problem of higher-order belief update in this context and presented some regularity results that connected beliefs (epistemology) and behavior (optimal strategies). Based on this, we developed a novel evenly-dispersed higher order belief sample generation scheme (the *spread-accumulate* method) for approximating higher-order belief identification in order to (approximately) realize higher-order belief update.

Our methods are potentially generalizable to other problem domains that involve strategic multiagent interactions – all that needs to be done is to check whether the epistemology-optimal behaviour regularity phenomenon holds for a given problem. A complete characterization of general epistemological and strategic conditions under which this phenomenon arises is crucial for advancing the finite epistemological decision-theoretic framework and for completing the theory of intelligent and autonomous behavior in multiagent settings.

REFERENCES

Aumann, R. and Brandenburger, A. (1995). Epistemic conditions for nash equilibrium. *Econometrica*, 63(5):1161–1180.

Banks, J. and Sobel, J. (1987). Equilibrium selection in signaling games. *Econometrica*, 55(3):647–661.

Cho, I.-K. (1987). A refinement of sequential equilibrium. *Econometrica*, 55(6):1367–1389.

Cho, I.-K. and Kreps, D. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 102(2):179–222.

Doshi, P. and Gmytrasiewicz, P. (2005). Approximating state estimation in multiagent settings using particle filters. In *In Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*.

Fudenberg, D. and Levine, D. (1981). Perfect equilibria of finite and infinite horizon games.

Gmytrasiewicz, P. and Doshi, P. (2005). A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, 24:49–79.

Harsanyi, J. C. (1968). Games with incomplete information played by 'bayesian' players, i-iii. *Management Science*, 14:pp. 159–182, 320–334, 486–502.

Kreps, D. (1985). Signaling games and stable equilibria.

Kreps, D. and Wilson, R. (1982). Sequential equilibria. *Econometrica*, 50:863–894.

Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49.

Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295.

Samuelson, W. F. (1984). Bargaining under asymmetric information. *Econometrica*, 52(4).

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55.

Sobel, J. and Takahashi, I. (1983). A multistage model of bargaining. *Review of Economic Studies*, 50(3).