Language-based Games

Adam Bjorndahl, Joseph Y. Halpern, and Rafael Pass

Cornell University

Introduction

In standard game theory, utility depends only on outcomes. But this does not account for many features of a situation that determine how an agent feels:

- beliefs
- expectations
- feelings such as guilt and surprise.

We allow utility to depend on descriptions in some *language*. This generalizes

- standard game theory: the language describes only the strategies used;
- psychological games [GPS '89; BD '09]: the language is extended to allow beliefs;
- reference-dependent preferences [KR '06]: the language can talk about expected outcomes as well as actual outcomes.

What's different?

- Considering a *coarse* language, where there are fewer descriptions than there are actual differences to describe, offers insights into a variety of puzzles and paradoxes of human decision making.
  - It also leads to technical differences: standard game theory tends to focus on *continuous* settings; with a language, things are typically *discrete*.
- The structure of the underlying language can determine whether certain kinds of equilibria exist.

A surprise proposal

Bob wants to propose to Alice, but he also wants it to be a surprise. In fact, if Alice expects the proposal, Bob would prefer to postpone it entirely; otherwise, if Alice is not expecting it, Bob's preference is to take the opportunity.

	B proposes (p)	B doesn't propose ($\neg p$)
A expects it ($B_A p$)	bad (0)	good (1)
A doesn't expect it ($\neg B_A p$)	good (1)	bad (0)

- Note that Bob's utility depends only on the truth of formulas in the language $\{p, \neg p, B_A p, \neg B_A p\}$.
- In general, the set of formulas that a player's utility can depend on is called the *underlying language*.

Situations and $\mathcal{L}$-games

- Let $\mathcal{L}$ be a *language*, that is, a set of well-formed formulas.
  - $\mathcal{L}$ can be finite or infinite.
  - In the surprise proposal, can take $\mathcal{L} = \{p, \neg p, B_A p, \neg B_A p\}$.
- An *$\mathcal{L}$-situation* is a maximal satisfiable set of formulas from $\mathcal{L}$.
  - Intuitively, it is a complete description of the world in the language $\mathcal{L}$.
- $\mathcal{S}(\mathcal{L})$ denotes the set of all $\mathcal{L}$-situations.
- Given a language $\mathcal{L}$, an *$\mathcal{L}$-game* has utility functions

$$u_i : \mathcal{S}(\mathcal{L}) \to \mathbb{R},$$

  for each player i.
- Thus, in an $\mathcal{L}$-game, a player's utility can depend on anything expressible in the underlying language $\mathcal{L}$, and nothing more.

Classical game theory as a special case

- A *game form* $\Gamma = (N, (\Sigma_i)_{i \in N})$ is a set of players N and, for each player $i \in N$, the set of their available strategies, Σ_i.
- Consider a propositional language $\mathcal{L}(\Phi_\Gamma)$ with primitive propositions $\Phi_\Gamma = \{play_i(\sigma_i) \,:\, i \in N, \sigma_i \in \Sigma_i\}$.
  - $play_i(\sigma_i)$ is read "player i is playing strategy σ_i".
- Given the obvious semantic constraints (each player plays exactly one strategy), maximal satisfiable sets of formulas from $\mathcal{L}(\Phi_\Gamma)$ correspond exactly to *strategy profiles*:
  - a description of what each player plays.
- A utility function defined on $\mathcal{L}(\Phi_\Gamma)$-situations is therefore equivalent to one defined on outcomes.
- A standard game based on Γ is thus realized as an $\mathcal{L}(\Phi_\Gamma)$-game.

Belief-dependent preferences

- Consider the language $\mathcal{L}_B(\Phi_\Gamma)$ generated by

$$\varphi \ ::= \ p \,|\, \neg\varphi \,|\, \varphi_1 \wedge \varphi_2 \,|\, B_i\varphi,$$

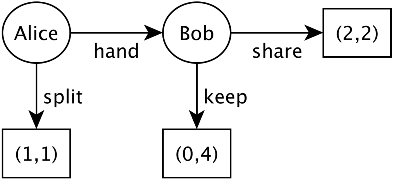
  where $p \in \Phi_\Gamma$ and $i \in N$.
  - $B_i\varphi$ is read "player i believes φ".
- This language can express what each player believes about which strategies are played (since it includes the formulas $play_i(\sigma_i)$), what each player believes about what their opponents believe, and so on.
- An $\mathcal{L}_B(\Phi_\Gamma)$-game is one in which the players' preferences can depend on beliefs as expressed by the belief modalities B_i.

To determine $\mathcal{L}_B(\Phi_\Gamma)$-situations (maximal satisfiable sets of formulas), we need to provide formal semantics (i.e., a notion of *truth*) for this language.

- For this we use a standard approach: possible worlds equipped with probability measures that interpret the belief modalities.
- $B_i\varphi$ is interpreted as full (i.e. probability 1) belief in φ.
- The formal details do not matter here.

The psychological trust game [Battigalli-Duwfenberg '09]

Alice can either split \$2 with Bob, or hand him all of it. If she splits the money, the game is over and they each keep \$1. If she hands the money to Bob, it is doubled to \$4. Bob can then either share the money with Alice, or keep it all for himself. However, if Bob chooses to keep the money for himself, then he suffers from guilt to the extent that he let Alice down.



We can capture Bob's guilt aversion by setting:

$$u_B(S) = \begin{cases} -1 & \text{if } \mathit{play}(\textbf{hand, keep}) \in S \\ & \text{and } B_A\, \mathit{play}_B(\textbf{share}) \in S \\ \text{payoffs above} & \text{otherwise.} \end{cases}$$

The trust game

A more satisfying account of this game might involve more than a binary representation of Alice's expectations.

- In psychological game theory, Bob's preferences would depend on Alice's beliefs in a "continuous" way:
  - If $\alpha \in [0, 1]$ is the probability Alice assigns to Bob playing **share**, we might define u'_B(**hand, keep**) $= 4 - 5\alpha$.
- A language-based game with an appropriately expressive underlying language can capture this, too.
  - E.g., a language with belief modalities B_i^{α} for each $\alpha \in [0, 1]$ read "player i considers the probability of φ to be α".
- However, there is good reason to think that human reasoning is sometimes better modeled by a coarser, categorical representation of belief [Mullainathan '02].
  - We might instead use, for example, a language with belief modalities B_i^{ℓ} for each integer $\ell \in [1, k]$.
  - Intuitively, ℓ ranges over discrete *levels of belief*: "unlikely", "somewhat likely", "very likely", etc.

Coarseness

Coarseness offers insight in a wide variety of contexts.

- Consumers often evaluate prices in a discontinuous way, behaving, for instance, as if the difference between \$2.99 and \$3.00 is more substantive than that between \$3.00 and \$3.99.
- This behaviour can be accounted for by dropping the implicit assumption that utility is a continuous function of cost.
  - Perhaps, in a typical shopping context, consumers do not regard \$3.00 and \$3.01 as distinct expenses.
    - We can capture this in the language-based framework by assuming that the language can say only "between \$2.00 and \$2.99", "between \$3.00 and \$3.99", etc.
    - Both \$3.00 and \$3.01 are represented as "between \$3.00 and \$3.01".
    - The agent certainly understands that \$3.00 and \$3.01 are different. However, these differences have no impact on utility.

Why do agents use coarse languages?

- Conjecture: they're cognitively easier to deal with

Cars vs. iPods

Suppose that an agent is about to buy a car for \$20,000, and hears that the identical car is available 3 miles down the road for \$100 less.

- \$100 is noise! The agent doesn't go anywhere.

Suppose that an agent is about to buy an iPod for \$200 and hears that the identical iPod is available 3 miles down the road for \$100 less.

- I'm outta here!

In our framework, there is a simple explanation:

- The range of prices subsumed under a single description is proportional to the magnitude of the price.
- Weber's Law: the size of a "just noticeable" difference is proportional to the magnitude of the stimulus.

The Allais paradox I

Coarseness can also explain the Allais paradox.

Gamble 1a		Gamble 1b	
1	$1 million	.89	$1 million
		.1	$5 million
		.01	$0

Gamble 2a		Gamble 2b	
.89	$0	.9	$0
.11	$1 million	.1	$5 million

- Most people prefer gamble 1a to gamble 1b, and prefer gamble 2b to gamble 2a.
- However, such preferences are not simultaneously compatible with expected utility maximization, no matter what utility is assigned to $0, $1M, and $5M.

The Allais paradox II

Suppose we were to represent these gambles using a language that can express only a finite number of distinct likelihoods:

True likelihood	Description	Approximation
1	certain	1
$[.95, 1)$	near certain	$.975$
$[.85, .95)$	likely	$.9$
$(.05, .15]$	unlikely	$.1$
$(0, .05]$	slight chance	$.025$
0	no chance	0

- The true likelihood is translated into the appropriate description in the language.
- Each description is approximated by a fixed probability value for the purposes of making calculations.

The Allais paradox III

Gamble 1a		Gamble 1b	
certain	\$1 million	likely	\$1 million
		unlikely	\$5 million
		slight chance	\$0

Gamble 2a		Gamble 2b	
likely	\$0	likely	\$0
unlikely	\$1 million	unlikely	\$5 million

- The distinction between $.1$ and $.11$, and between $.89$ and $.9$, disappears with this language.
- The distinction between 0 and $.01$, however, is preserved.
- With these descriptions, preferring 1a to 1b and 2b to 2a *is* compatible with expected utility maximization.

Reference-dependent preferences

Alice can give Bob a salary increase somewhere between \$0 and \$1/hour. Bob's happiness is determined in part by the raw value of the bump he receives in his wages, and in part by the sense of gain or loss he feels by comparing the increase Alice grants him with the minimum increase he expected to get. Alice wants Bob to be happy, but also wants to save money.

Alice can play $s_0, \ldots, s_{100}$, where s_k gives Bob raise of k cents/hr.

- $u_B(S) = k + f(k - r)$, where
  - k is the unique integer such that $play_A(s_k) \in S$: Alice's action in situation S;
  - $r = \min\{r' : \neg B \neg play_A(s_{r'}) \in S\}$:
    - r is the smallest raise that Bob assigns positive probability to
  - f determines how Bob's utility is affected by his expectations. Could have, e.g., $f(x) = \alpha x$ if $x \geq 0$; $f(x) = \beta x$ if $x < 0$.
- $u_A(S) = -|k - r|$.
  - Alice wants to give Bob what he expects, no more and no less.

Intentions

Suppose Alice has to make a decision today, but how she feels today depends on what she might do tomorrow. For example:

- exercising
- quitting smoking.

Simplified version: Alice can return a library book today (**return**) or wait (**wait**). She must return it by tomorrow (captured by the proposition tomorrow) to avoid a fine. Define

$$u_A(S) = \begin{cases} -1 & \text{if } play_A(\mathbf{return}) \in S \\ 1 & \text{if } play_A(\mathbf{wait}) \wedge B_A \mathsf{tomorrow} \in S \\ -5 & \text{otherwise;} \end{cases}$$

Alice prefers to wait if she expects to return the book tomorrow, and to return the book today otherwise.

Alice might even take actions to ensure that she really returns the book tomorrow . . .

Rationality

What does it mean for a player to be *rational* in this framework?

In standard game theory, we say that player i is rational at a state ω where i is using strategy σ_i if i's expected utility at ω is at least as high as it would be if i switched to some other strategy σ_i'.

To make sense of this, we must assume

- associated with each state ω is a strategy for each player, and a distribution for player i on states.
  - This induces a distribution on what other players are playing.
  - This lets i compute his expected utility at ω.

But what is i's expected utility if he switches to σ_i'?

- This can be computed thanks to a standard assumption:
  - i's beliefs about what other players do remain the same
    - (*not* assumed in [Halpern-Pass '13]).

Rationality in our framework

In our framework, utility is assigned to situations, so i's expected utility depends on i's beliefs about other situations.

- Each state uniquely determines a situation, so a distribution on states allows us to calculate expected utility, as before.

But rationality means maximizing expected utility, so we also need to calculate i's expected utility if he switches to σ_i'.

- Switching to σ_i' changes i's beliefs.
- We can define "the situation S' that is just like S except that i switched to σ_i'", and so define rationality.
  - But that depends on the fact that $\mathcal{L}_B(\Phi_\Gamma)$ talks only about beliefs about strategies.
  - In general, to define rationality in S, we need some way of defining the situation S' that would result if i switched to σ'.

Solution concepts

How should we expect an agent to play in a game?

- *Solution concepts* such as Nash equilibrium make predictions in standard games.
- Are there analogues in our setting?

In the standard setting, $(\sigma_1, \ldots, \sigma_n)$ is a *Nash equilibrium* if, for all players i and all strategies σ_i' for i:

$$EU_i(\sigma_1, \ldots, \sigma_n) \geq EU_i(\sigma_i', \sigma_{-i}).$$

- Gloss: i is happy with his strategy *even knowing everyone else's strategy*.

Nash equilibrium may not exist in our setting ...

Indignant altruism I

Alice and Bob play a standard game of prisoner's dilemma, with one twist: neither wishes to live up to low expectations. If Bob expects the worst of Alice (i.e. expects her to defect), then Alice, indignant, prefers to cooperate. Likewise for Bob.

Define $u_A : \mathcal{S}(\mathcal{L}_B(\Phi_\Gamma)) \to \mathbb{R}$ by

$$u_A(S) = \begin{cases} -1 & \text{if } \mathit{play}_A(\mathbf{d}) \wedge B_B\, \mathit{play}_A(\mathbf{d}) \in S \\ \text{“standard” payoffs} & \text{otherwise.} \end{cases}$$

and similarly for u_B, where the "standard" payoffs are as follows:

	c	**d**
c	(3,3)	(0,5)
d	(5,0)	(1,1)

Indignant altruism II

Intuitively, this game admits no Nash equilibrium (pure or mixed).

- Cooperating is rational for Alice if she thinks that Bob is sure she'll defect.
- On the other hand, if Alice thinks that Bob is *not* sure she'll defect, then it is rational for her to defect.
- But a Nash equilibrium, intuitively, is a state of play where everyone is happy with their choice of strategy *even having correct beliefs about what their opponents will do*.
    - There is a fundamental tension between a state of play where everyone has correct beliefs, and one where some player successfully surprises another.

Of course, to formalize this argument, we need to actually *define* Nash equilibrium in $\mathcal{L}_B(\Phi_\Gamma)$-games.

Epistemic characterizations

There is a natural way to generalize standard solution concepts.

- Nash equilibrium has been characterized epistemically: a mixed strategy profile σ is a Nash equilibrium precisely when the state space M_σ corresponding to everone playing σ satisfies common belief of rationality.
  - At each state in M_σ, player i's beliefs about the other players' strategies is determined by σ_{-i}
  - It is immediate that this generalizes the standard definition.
- But Nash equilibrium has other epistemic characterizations [Aumann-Brandenburger '95]
  - As long as certain minimal properties are required, there is no Nash equilibrium in the indignant altruism game.

Any solution concept with an epistemic characterization can similarly imported into the language-based framework.

- Indignant altruism also has no *correlated equilibrium*.

Rationalizability [Bernheim/Pearce, 1984]

Another standard solution concept: *rationalizability*.

- More general than Nash: every strategy in a Nash equilibrium is rationalizable.
- Rationalizable strategies can be defined as the outcome of an iterative deletion procedure.
- Epistemic characterization [Tan-Werlang, 1988; Brandenburger-Dekel 1987]: σ_i is rationalizable iff it can be played in a state where rationality is common belief.

In the language-based framework, we can use the epistemic characterization as the definition of rationalizability.

- Every strategy in the indignant altruism game is rationalizable.
- There exist games with no rationalizable strategy.
- Every $\mathcal{L}_B(\Phi_\Gamma)$-game Γ satisfying a weak condition has rationalizable strategies:
  - Roughly, an agent's irrationality in a situation must be determined by finitely many formulas.
  - The proof uses an iterative deletion procedure, and relies on the *compactness* of $\mathcal{L}_B$.

Summary

We have defined a language-based framework for games.

- *Outcomes* are replaced by *situations* as the objects over which players form preferences.
  - The expressive power of the underlying language directly determines the kind of preferences players may form.
- Psychological effects can be modeled with an underlying language that expresses beliefs.
- Coarse beliefs are naturally represented and prove to be a useful modeling tool.
- For $\mathcal{L}_B(\Phi_\Gamma)$-games, Nash equilibria do not exist in general, but rationalizable strategies do, under mild assumptions.
  - The proofs show the importance of the structure of the underlying language.
  - They also show that meaningful solution concepts can be defined in a non-continuous setting.

Further research

The language-based framework arguably gives us too much flexibility. What are natural language assumptions and their impact?

- Coarseness:
  - special role of 0;
  - small quantities all seem the same (10^{-15} looks like 10^{-9});
  - Weber's Law;
  - the choice of "buckets" and anchoring effects;
  - coarse representations of uncertainty.
- The role of language in the formation of preferences:
  - using "intention" to capture phenonmena like procrastination;
  - temporal logics for modelling preferences in dynamic games;
  - using awareness for modelling players who care about what language their opponents reason with.

We have focused on a static framework. What about dynamics?

- Advertisers try to get you to change your language!