Formal Models: Section 3*

Zachary Lorico Hertz

2 February 2026

Review: from last week

Recall from our first week that the central actors in this course are decision-makers who select from a set of *actions* and hold preferences over *outcomes*. In our last section, we faced an important question: how do decision-makers evaluate preferences when outcomes are uncertain?

To answer this question, we reviewed the concept of **expected utility**. In response to uncertainty, decision-makers consider how much utility they would receive from particular outcomes, weight that outcome by the probability that it will be realized, and then assess their decisions given the total distribution of these possibilities. We noted that the **utility function** itself not only determines how many "jollies" decision-makers receive from particular outcomes, but also to what extent they are willing to trade certain gain for uncertain gain (ie. their "risk tolerance").

When modeling, the final quantity we are often interested in calculating is the *probability that a decision-maker selects a particular action.* To recover this, we can take statements of the rules actors use to determine what decision they select; modify these statements so that they are about the purely stochastic component in comparison to the purely deterministic component;[1] use properties of probability distributions to make a statement expressed as the probability that an action will be selected; and, when necessary, use derivatives to determine how this probability changes based on its inputs.[2]

Important properties of CDFs, more generally

At this point, you should have more or less all of the tools you'll need to approach Problem Set 1 with grit and vigor. Last week's handout focused mostly on CDF properties for the special kind of

*I drew on materials from Tak-Huen Chau, who himself drew on materials from David Foster and Jieun Kim. All writing (including its poor sense of humor and any errant mistakes) are mine.

[1] Sorry for the horribly technical sentence, which I'm keeping as-is for brevity. What this is saying is that we want to use algebra or mathematical manipulation to write the statement as a comparison between the parts that are probabilistic and change and the parts that are unchanging at the individual level. In our exercise 2 from last week, for example, we isolated ϵ, which was probabilisticly distributed, from the benefits β and costs C that were static.

[2] An anonymous index card requested more explanation on how we move from utility functions to the probability distributions. If this was you and you feel like more explanation is needed, please come to office hours or reach out again!

distribution (mean-zero, symmetric, single-peaked) that you'll encounter. I want to discuss more general properties that may also be useful to you as you work through these problems formally.

Recall that for a given random variable X, we can define the cumulative distribution function (CDF) of X as F_X. Then, the probability that the variable X is less than or equal to some arbitrary value x is written as $F_X(x) = \mathbb{P}(X \leq x)$.[3] But what if we are interested in the probability that X is *greater* than some value?

To find this will require a bit of a detour. Recall also from math camp that any event has a counterpart that we call its *complement*. The *complement* of event E is the event [Not E]. For any event E, we can also say that the probability of its complement occurring, or $\mathbb{P}(\text{Not } E)$, is equal to 1 minus the probability that event occurs $(1 - \mathbb{P}(E))$.

To illustrate this, imagine I am playing Texas hold 'em poker, a card game where each player is initially dealt two cards out of a set of 52. In this game receiving two aces from the dealer is a very powerful set of cards, which we call "pocket aces," and we might imagine that I am interested to know what the probability is of this event, which we will call the event E. I'll drop the actual proof to a footnote[4] and tell you up front that the probability of getting pocket aces in Texas Hold 'Em is $\frac{1}{221}$. Then, in this example the *complement* event is the one where I am not dealt pocket aces, and the probability that I am not dealt pocket aces can be simply calculated as $1 - \frac{1}{221} \approx 99.5\%$.

Given this rule, we can establish **the probability that X is *greater* than some value**:

$$\begin{aligned} \mathbb{P}(X > x) &= 1 - \mathbb{P}(X \leq x) && \text{(By definition of complement)} \\ &= 1 - F_X(x) && \text{(By definition of CDF)} \end{aligned}$$

We might also be interested in **the probability that X lies in some interval** $(a, b]$, ie. is *greater* than some value a and less than or equal to some value b. To do this, we take advantage of the fact that CDFs are **monotonic**, which is to say that if $a \leq b$, then $F_X(a) \leq F_X(b)$. As a result, we can write the interval $(a, b]$ as equal to the difference between the interval $(-\infty, b]$ and $(-\infty, a]$:

$$\begin{aligned} \Pr(a < X \leq b) &= \Pr(X \leq b) - \Pr(X \leq a) && \text{(By partitioning the event)} \\ &= F_X(b) - F_X(a) && \text{(By definition of CDF)} \end{aligned}$$

Finally, we might be looking at CDFs for continuous random variables which are *not* centered around 0, but instead have some other mean μ. For these random variables, we can state more generally:

$$F_X(\mu - x) = 1 - F_X(\mu + x) \forall x \in \mathbb{R}$$

[3]In section and lecture we may drop the subscript for brevity and write it as $F(x)$. But don't forget that we are really referring to $F_X(x)$; if we were using multiple CDFs it would be important to denote that we are specifically referring to the CDF of X with F_X.

[4]The probability of being dealt pocket aces is $\frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221} \approx 0.452\%$. This comes from the fact that there are 4 aces in a standard 52-card deck, so the first card has a $\frac{4}{52}$ chance of being an ace, and then given that the first card is an ace, there are 3 aces remaining out of 51 cards, giving a $\frac{3}{51}$ probability for the second card.

“I want to play a game.”

Hopefully this statement inspires less fear in this context than it does in the *Saw* franchise. Nevertheless, this week we’re finally putting the “game” in “game theory”! The rest of this handout is devoted to a review of preliminary information in our introduction to **strategic form games**.

To this date, our framework could be summarized succinctly as the punchy reminder “actions map to outcomes, preferences map to utilities.” In slightly longer form, we’ve previously discussed models in which decision-makers, and in particular an individual i chooses an action from a set of possible actions A. There are also a set of outcomes X, some mapping through which actions relate to outcomes that we call $\mu : A \to X$, and the decision-maker’s preferences over the possible outcomes, which we represent with the utility function $u : x \to \mathbb{R}$.

As-is, this framework has plenty of advantages. At first glance, it has many of the traits we seek in formal analysis: it is precise, it provides a way to mathematically recover predictions from preferences, and it can be applied to a wide variety of contexts. There is one implication of this model, however, that is somewhat shaky ground for many types of theories we are interested in. Specifically, in this framework, by assumption we have defined outcomes as consequences *of, and **only of** the decision-maker’s actions in combination with some randomness beyond their control.* As a result, in this model, when choosing their action, the decision-maker that we have discussed cares only about their selected action, insofar as it translates to outcomes.

When setting out original models, we ought to consider two key questions: “Is this model tractable?”, and “Does this model reflect the dynamics we observe in the real world?”. With these questions in mind, one potential critique of our models to this point may surface: the assumption that decision-makers choose actions unaffected by the actions of others seems unlikely to accurately model the world we live in. Put another way, sometimes your outcomes **are** dependent on the choices made by other decision-makers.

Consider, for example, a situation where my office-mate Andrew and I agree to get lunch on an agreed-upon date. In the event I return to this example in the future, I name this example **“Battle of the Cohort I.”** When the time for lunch arises, however, Andrew and I realize that we neglected to agree upon a restaurant. As graduate students with busy schedules, we find ourselves under time and material budget constraints, so a place on Telegraph seems suitable. But we also both know that we have vastly different preferences. I prefer Mezzo, but Andrew finds their giant sandwiches and bountiful salads to be far too much food for a single meal. Andrew prefers Sinaloa, but I think their food is greasy enough to ruin an afternoon. There are stakes, and thus there is conflict! So, we decide to settle the issue with a single game of rock paper scissors[5] where the winner gets to choose the restaurant; in the event of a tie, we go to the mutual compromise restaurant D’Yar.

Hopefully, at this point, it should be clear that my outcome, my lunch, and my very well-being depend not only on my choice in the rock paper scissors showdown, but also on Andrew’s choice. A misinformed purist of the previous framework might protest that we should just model Andrew’s choice as part of the randomness of the universe that lies beyond my control. But recall our model of rational choice would not find this satisfactory. Andrew is not going to pick his choice randomly; he *also* has stakes in the outcome of this lunch, and thus should try to determine what I play in order to decide what his best pick is. We characterize these situations, where actors must assess other actors’ decisions, as a *strategic environment.*

[5]Known alternatively as *Rochambeau*, or if you are Filipino like I am, “bato bato pik”.

This reveals a second shortcoming of our previous approach; if we do acknowledge that the actions of other decision-makers become variables that affect the action of some decision-maker i, recovering a tractable solution can quickly become quite complicated. We start with the simplest language to represent these strategic situations, which we call **games**. Because we are starting with the simplest examples, we make a key assumption: that these games are *static*.

Definition (Static): A game is *static* if players independently choose their actions, which cannot be changed and are immediately realized, after which outcomes are distributed conditional on the combination of players' choices.

Static games can be considered to be a two-step process: in the first, players *simultaneously and independently* choose their actions. This does not necessarily mean that players choose at literally the exact same moment in time; what it does require is that players cannot observe the actions of other players before deciding their actions. They cannot coordinate, nor do they interact before deciding. In the second, payoffs are distributed conditional on the combination of players' selected choices; in other words, the choices jointly determine what outcome we reach. In "Battle of the Cohort I" both my choice of rock, paper or scissors *and* Andrew's choice determine the winner of the game, and thus the restaurant we choose.

In our introduction to games, we start with the most fundamental type: **strategic-form games of complete information.** Note that this definition may use some terms that you are currently unfamiliar with! Don't worry - we'll get there, but for now we start with the formal definition of strategic-form games, and work backwards:

Definition (Strategic-form games): A *strategic-form game* is a general framework to represent games of the most basic form. They are characterized by three components:

1. There are a finite **set of players**, $N = \{1, 2, ..., n\}$ in the game.
2. There is a collection of **sets of pure strategies** for each player i in the game, $S = \{S_1, S_2, ..., S_n\}$. Each set of pure strategies is further defined as a collection of m finite strategies for player i such that $S_i = \{s_i^1, s_i^2, ..., s_i^m\}$.
3. There is a **set of utility functions** for each player in the game, $V = \{v_1, v_2, ..., v_n\}$ that assigns a payoff value to every possible strategy profile.

Defining each characteristic of strategic-form games

As promised, we now begin to work backwards through this definition. The first premise is straightforward: there is a countable, finite collection of players in the game. Each player i is assigned a number from 1 to n. We consider each of these players to be rational actors who have their own preferences over outcomes. Recall that before we had *actions* that mapped to *outcomes*; a *strategy* is a development of this idea.

Definition (Strategy): A *strategy* is often defined[6] as a plan of action intended by an actor to accomplish some specific goal.

For now, in the interest of simplicity we are interested in a particular type of strategy called a **pure strategy**:

Definition (Pure strategy): A *pure strategy* for player i is a deterministic plan of action.

[6] Indeed, here I borrow directly from Tadelis.

This is equivalent to the set of actions that each player can take. When we refer to these strategies as "pure," what we mean is that players choose a *certain* plan of action. In "Battle of the Cohort I," we can say that I have a set of pure strategies S_Z which can be represented as $S_Z = \{\text{Rock}, \text{Paper}, \text{Scissors}\}$. We refer to the element Rock as a pure strategy because if I pick the action Rock, I will definitely play Rock. This is in contrast with a *stochastic* strategy, in which I could decide that I will leave it to chance and roll a dice and throw a sign corresponding to the number on the die: Rock if the die shows 1 or 2, Paper if the die shows 3 or 4, and Scissors if the die shows 5 or 6. We can see, then, that this game has a collection of sets of pure strategies for both players: $S = \{S_Z, S_A\}$, and the second premise is satisfied.[7] To relate how these sets of strategies interact, we use **strategy profiles**.

Definition (Strategy profile): A *strategy profile* describes some combination of pure strategies for every player in the game. Usually, for players $N = \{1, 2, ...n\}$ with pure strategies $s_i \in S_i$, we refer to strategy profiles with the form $s = (s_1, s_2, ...s_n)$.

We defined $S_Z = \{\text{Rock}, \text{Paper}, \text{Scissors}\}$, and in this simple example the other player faces a similar set of choices $S_A = \{\text{Rock}, \text{Paper}, \text{Scissors}\}$. The collection of all strategy profiles in this game is just the Cartesian product of S_Z and S_A,[8] so we can define the total set of pure strategy profiles in this game as the set:

$$\begin{aligned} S = \{&(\text{Rock}_Z, \text{Rock}_A), (\text{Rock}_Z, \text{Paper}_A), (\text{Rock}_Z, \text{Scissors}_A), \\ &(\text{Paper}_Z, \text{Rock}_A), (\text{Paper}_Z, \text{Paper}_A), (\text{Paper}_Z, \text{Scissors}_A), \\ &(\text{Scissors}_Z, \text{Rock}_A), (\text{Scissors}_Z, \text{Paper}_A), (\text{Scissors}_Z, \text{Scissors}_A)\} \end{aligned}$$

Recall that our third premise states that every possible strategy profile must correspond to a payoff value for every player in the game. In strategic-form games, the strategy profile determines the realized outcome. By the rules we set above, we can see that each strategy corresponds to the following outcomes:

	Rock_A	Paper_A	Scissors_A
Rock_Z	Tie - D'Yar	Andrew wins - Sinaloa	Zach wins - Mezzo
Paper_Z	Zach wins - Mezzo	Tie - D'Yar	Andrew wins - Sinaloa
Scissors_Z	Andrew wins - Sinaloa	Zach wins - Mezzo	Tie - D'Yar

We take as an assumption the fact that Andrew and I have equally strong preferences for our preferred restaurant, and gain 3 jollies from eating at the restaurant we prefer. We have mild preferences for our less-preferred compromise, and arbitrarily state that we both gain 1 jolly from eating at D'Yar. Finally, I state that we both have negative utility from eating at our least-preferred

[7] At this point, I'd stop and ask yourself how we would write the strategy set for Andrew S_A. See how we are using the language we learned in Section 1?

[8] Again, refer to the Section 1 handout if you need to review how to calculate this. We are building on our previous material!

restaurant, which yields a payoff of -1 jolly. So, we can write the payoffs as follows:

$$v_Z(\text{Rock}_Z, \text{Scissors}_A) = v_Z(\text{Paper}_Z, \text{Rock}_A) = v_Z(\text{Scissors}_Z, \text{Paper}_A) = 3$$
$$v_A(\text{Scissors}_Z, \text{Rock}_A) = v_A(\text{Rock}_Z, \text{Paper}_A) = v_A(\text{Paper}_Z, \text{Scissors}_A) = 3$$
$$v_Z(\text{Scissors}_Z, \text{Rock}_A) = v_Z(\text{Rock}_Z, \text{Paper}_A) = v_Z(\text{Paper}_Z, \text{Scissors}_A) = -1$$
$$v_A(\text{Rock}_Z, \text{Scissors}_A) = v_A(\text{Paper}_Z, \text{Rock}_A) = v_A(\text{Scissors}_Z, \text{Paper}_A) = -1$$
$$v_Z(\text{Rock}_Z, \text{Rock}_A) = v_Z(\text{Paper}_Z, \text{Paper}_A) = v_Z(\text{Scissors}_Z, \text{Scissors}_A) = 1$$
$$v_A(\text{Rock}_Z, \text{Rock}_A) = v_A(\text{Paper}_Z, \text{Paper}_A) = v_A(\text{Scissors}_Z, \text{Scissors}_A) = 1$$

Putting this all together, we represent the game in a matrix with payoffs. These generally are written as a tuple of each player's payoffs in order, of the form $P = (p_1, p_2, ...p_n)$:

	Rock_A	Paper_A	Scissors_A
Rock_Z	(1,1)	(-1,3)	(3,-1)
Paper_Z	(3,-1)	(1,1)	(-1,3)
Scissors_Z	(-1,3)	(3,-1)	(1,1)

Knowing is half the battle

Recall that the second characteristic of these most fundamental games is that they are games **of complete information**. This means that there are four components, which are all *common knowledge*:

Definition (Complete information): A game of *complete information* is one in which all players of the game have common knowledge of each of these four components:

1. all possible actions of all the players, ie. $A_1, A_2, ...A_n$.
2. all possible outcomes, X.
3. how strategy profiles determine outcomes, ie. some function $S_1 \times S_2 \times ... \times S_n \to X$ that maps each strategy profile $s_k = (s_1, s_2, ...s_n)$ to a realized outcome $x \in X$ and
4. the preferences of each and every player over outcomes, ie. each utility function $v_i \in V$.

Returning to our two key questions: "Is this model tractable?", and "Does this model reflect the dynamics we observe in the real world?", we admit that the assumptions of complete assumption are so strong as to likely fail the latter question in many cases. But, as in our original definition of rational choice we turn to them because they certainly answer the former and provide a great deal of explanatory power over many cases. Finally, you may notice that I snuck in a final term that may be unfamiliar to you: *common knowledge.* This is our last demanding assumption for tractability over simple strategic-form games.

Definition (Common knowledge): An event E is *common knowledge* if (1) everyone knows E, (2) everyone knows that everyone knows E, and so on *ad infinitum.*

To illustrate why requiring common knowledge matters, consider a small high school where the principal, Dr. Em Bezler, has provided a case study in nominanative determinism and been embezzling funds. Suppose several teachers independently discover evidence and quietly share their suspicions with individual colleagues. Many teachers might know about the embezzlement, but this is not common knowledge. Teacher A doesn't know whether Teacher B has heard the rumors, Teacher B doesn't know that Teacher A knows, and so forth.

Now suppose instead that an auditor presents their findings at a mandatory all-staff meeting, announcing the embezzlement to the entire faculty at once. Not only does everyone know about the corruption, but everyone knows that everyone else was in the room and heard the announcement. Everyone knows that everyone knows that everyone knows, ad infinitum. In this scenario, the embezzlement is common knowledge.

Why does the distinction matter? In the first case, even if most teachers privately believe Dr. Bezler is corrupt, no individual teacher knows whether their colleagues share this belief, making coordinated action (like a collective complaint to the school board) risky to initiate. In the second case, the shared public nature of the announcement makes collective action far more feasible. We will return to this later, but *common knowledge* in strategic environments is incredibly powerful.

Summary

Last week someone requested a summary in 1-2 sentences. I'm afraid that's insufficient given the breadth of material, but I suppose I would say that **strategic-form games are the basic framework through which we model strategic interactions where outcomes depend on the choices of multiple rational actors. These depend on critical assumptions of complete information and common knowledge that ultimately enable us to predict decision-making.**

Less tersely, moving from individual decision-making to strategic interactions, we have introduced *games*, the fundamental framework we use in this course to analyze situations where outcomes depend on the choices of multiple actors. This is the "meat and potatoes" of this class, since so many political and social phenomena central to political science (voting, negotiations, conflict, collective action) focus on models of how rational actors respond to each other's decisions.

We start with **strategic-form games** as our basic tool for representing these interactions. A strategic-form game consists of three components: a set of players N, a collection of strategy sets S for each player, and utility functions V that assign payoffs to every possible strategy profile. We focus for now on *pure* strategies (deterministic plans), as opposed to *mixed* strategies (stochastic plans that randomize over actions), and define strategy profiles $s = (s_1, s_2, ..., s_n)$ to represent the combination of all players' choices.

We ended on a discussion of how the assumption of **complete information** is critical to analyzing these basic games. Under complete information all players have common knowledge of the available strategies, possible outcomes, how strategies map to outcomes, and each player's preferences. This turns out to be incredibly powerful in strategic environments because it enables coordination and collective action.

As a reminder, we're starting with the simplest case (static, complete information games) for tractability. These demanding assumptions give us explanatory power over many situations, and will serve as our baseline before we introduce complications like sequential moves, incomplete information, and repeated interactions.

Concepts covered:

- Strategic environments and static games

- Strategic-form games: players N, strategy sets S, utility functions V
- Pure strategies (deterministic) vs. mixed strategies (stochastic)
- Strategy profiles $s = (s_1, s_2, ..., s_n)$ and payoff matrices
- Complete information: common knowledge of strategies, outcomes, outcome function, and preferences
- Common knowledge and infinite regress (E, everyone knows E, everyone knows everyone knows E, etc.)