1
Introduction

First-order logic meets game theory as soon as one considers sentences with alternating quantifiers. Even the simplest alternating pattern illustrates this claim:

$$\forall x \exists y (x < y). \tag{1.1}$$

We can convince an imaginary opponent that this sentence is true on the natural numbers by pointing out that for every natural number m he chooses for x, we can find a natural number n for y that is greater than m. If, on the other hand, he were somehow able to produce a natural number for which we could not find a greater one, then the sentence would be false.

We can make a similar arrangement with our opponent if we play on any other structure. For example, if we only consider the Boolean values 0 and 1 ordered in their natural way, we would agree on a similar protocol for testing the sentence, except that each party would pick 0 or 1 instead of any natural number.

It is natural to think of these protocols as *games*. Given a first-order sentence such as (1.1), one player tries to verify the sentence by choosing a value of the existentially quantified variable y, while the other player attempts to falsify it by picking the value of the universally quantified variable x. Throughout this book we will invite Eloise to play the role of verifier and Abelard to play the role of falsifier.

We can formalize this game by drawing on the classical theory of extensive games. In this framework, the game between Abelard and Eloise that tests the truth of (1.1) is modeled as a two-stage game. First Abelard picks an object m. Then Eloise observes which object Abelard chose, and picks another object n. If $m < n$, we declare that Eloise has won the game; otherwise we declare Abelard the winner. We notice that

Eloise's ability to "see" the object m before she moves gives her an advantage. The reason we give Eloise this advantage is that the quantifier $\exists y$ lies within the scope of $\forall x$. In other words, the value of y depends on the value of x.

Hintikka used the game-theoretic interpretation of first-order logic to emphasize the distinction between constitutive rules and strategic principles [28, 29]. The former apply to individual moves, and determine whether a particular move is correct or incorrect. In other words, constitutive rules determine the set of all possible *plays*, i.e., the possible sequences of moves that might arise during the game. In contrast, strategic principles pertain to the observed behavior of the players over many plays of the game. Choosing blindly is one thing, following a strategy is another. A strategy is a rule that tells a particular player how to move in every position where it is that player's turn. A winning strategy is one that ensures a win for its owner, regardless of the behavior of the other player(s). Put another way, constitutive rules tell us how to play the game, while strategic principles tell us how to play the game well.

When working with extensive games, it is essential to distinguish between winning a single play, and having a winning strategy for the game. If we are trying to show that (1.1) holds, it is not enough to exhibit one single play in which $m = 4$ and $n = 7$. Rather, to show (1.1) is true, Eloise must have a strategy that produces an appropriate n for each value of m her opponent might choose. For instance, to verify (1.1) is true in the natural numbers, Eloise might use the winning strategy: if Abelard picks m, choose $n = m + 1$. If we restrict the choice to only Boolean values, however, Abelard has a winning strategy: he simply picks the value 1. Thus (1.1) is true in the natural numbers, but false if we restrict the choice to Boolean values.

To take an example from calculus, recall that a function f is *continuous* if for every x in its domain, and every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all y,

$$|x - y| < \delta \quad \text{implies} \quad |f(x) - f(y)| < \varepsilon.$$

This definition can be expressed using the quantifier pattern

$$\forall x \forall \varepsilon \exists \delta \forall y (\ldots), \tag{1.2}$$

where the dots stand for an appropriate first-order formula. Using the game-theoretic interpretation, (1.2) is true if for every x and ε chosen by Abelard, Eloise can pick a value for δ such that for every y chosen by Abelard it is the case that ...

The key feature of game-theoretic semantics is that it relates a central concept of logic (truth) to a central concept of game theory (winning strategy). Once the connection between logic and games has been made, logical principles such as bivalence and the law of excluded middle can be explained using results from game theory. To give one example, the principle of bivalence is an immediate consequence of the Gale-Stewart theorem, which says that in every game of a certain kind there is a player with a winning strategy.

Mathematical logicians have been using game-theoretic semantics implicitly for almost a century. The *Skolem form* of a first-order sentence is obtained by eliminating each existential quantifier, and substituting for the existentially quantified variable a *Skolem term* $f(y_1, \ldots, y_n)$, where f is a fresh function symbol and $y_1, \ldots, y_n$ are the variables upon which the choice of the existentially quantified variable depends. A first-order formula is true in a structure if and only if there are functions satisfying its Skolem form.

For instance the Skolem form of (1.1) is $\forall x\big(x < f(x)\big)$. In the natural numbers, we can take f to be defined by $f(x) = x + 1$, which shows that (1.1) is true. Thus we see that Skolem functions encode Eloise's strategies.

Logic with imperfect information

The game-theoretic perspective allows one to consider extensions of first-order logic that are not obvious otherwise. Independence-friendly logic, the subject of the present volume, is one such extension.

An extensive game with imperfect information is one in which a player may not "see" ("know") all the moves leading up to the current position. Imperfect information is a common phenomenon in card games such as bridge and poker, in which each player knows only the cards on the table and the cards she is holding in her hand.

In order to specify semantic games with imperfect information, the syntax of first-order logic can be extended with slashed sets of variables that indicate which past moves are unknown to the active player. For example, in the independence-friendly sentence

$$\forall x \forall y \big(\exists z / \{y\}\big) R(x, y, z), \tag{1.3}$$

the notation $/\{y\}$ indicates that Eloise is not allowed to see the value of y when choosing the value of z.

Imperfect information does not prevent Eloise from performing any particular action she could have taken in the game for the first-order variant of (1.3):

$$\forall x \forall y \exists z R(x, y, z). \tag{1.4}$$

Instead, restricting the information available to the player prevents them from following certain strategies. For instance, in the game for (1.4) played on the natural numbers, Eloise may follow the strategy that takes $z = x + y$. However, this strategy is not available to her in the game for (1.3).

The restriction on Eloise's possible strategies is encoded in the Skolem form of each sentence. For instance, the Skolem form of (1.3) is

$$\forall x \forall y R\big(x, y, f(x)\big),$$

whereas the Skolem form of (1.4) is

$$\forall x \forall y R\big(x, y, f(x, y)\big).$$

The set under the slash in $\big(\exists z/\{y\}\big)$ indicates that the quantifier is *independent* of the value of y, even though it occurs in the scope of $\forall y$.

Returning to calculus, a function f is *uniformly continuous* if for every x in its domain and every $\varepsilon > 0$, there exists a $\delta > 0$ independent of x such that for all y,

$$|x - y| < \delta \quad \text{implies} \quad |f(x) - f(y)| < \varepsilon.$$

The definition of uniform continuity can be captured by an independence-friendly sentence of the form

$$\forall x \forall \varepsilon \big(\exists \delta/\{x\}\big) \forall y (\ldots),$$

or, equivalently, by a first-order sentence of the form

$$\forall \varepsilon \exists \delta \forall x \forall y (\ldots).$$

Not all independence-friendly sentences are equivalent to a first-order sentence, however. Independence-friendly (IF) logic is related to an earlier attempt to generalize first-order logic made by Henkin [25], who introduced a two-dimensional notation called *branching quantifiers*. For instance, in the branching-quantifier sentence

$$\begin{pmatrix} \forall x & \exists y \\ \forall z & \exists w \end{pmatrix} R(x, y, z, w) \tag{1.5}$$

the value of y depends on x, while the value of w depends on z. The Skolem form of the above sentence is given by:

$$\forall x \forall z R\big(x, f(x), z, g(z)\big).$$

We can obtain the same Skolem form from the IF sentence

$$\forall x \exists y \forall z \big(\exists w / \{x, y\}\big) R(x, y, z, w). \tag{1.6}$$

Ehrenfeucht showed that sentences such as (1.5) can define properties that are not expressible in first-order logic [25]. Since branching-quantifier sentences are translatable into IF sentences, IF languages are also more expressive than first-order languages. In fact, IF logic has the same expressive power as existential second-order logic.

The additional expressive power of independence-friendly logic was the main reason why Hintikka advocated its superiority over first-order logic for the foundations of mathematics [28].

Several familiar properties of first-order logic are lost when passing from perfect to imperfect information. They will be discussed in due time. Here we shall briefly consider two such properties. It will be seen that the Gale-Stewart theorem fails for extensive games with imperfect information, and thus there is no guarantee that every IF sentence is either true or false.

One such notorious IF sentence is

$$\forall x \big(\exists y / \{x\}\big) x = y. \tag{1.7}$$

Even on a small domain like the set of Boolean values, Eloise has no way to consistently replicate the choice of Abelard if she is not allowed to see it. Abelard does not have a winning strategy either, though, because Eloise may guess correctly.

Thus, allowing semantic games of imperfect information introduces a third value in addition to true and false. It has been shown that the propositional logic underlying IF logic is precisely Kleene's strong, three-valued logic [31, 34].

Another familiar property of first-order logic that is often taken for granted is that whether an assignment satisfies a formula depends only on the values the assignment gives to the free variables of the formula. In contrast, the meaning of an IF formula can be affected by values assigned to variables that do not occur in the formula at all. This is exemplified by sentences such as

$$\forall x \exists z \big(\exists y / \{x\}\big) x = y. \tag{1.8}$$

In the semantic game for the above sentence, Eloise can circumvent the informational restrictions imposed on the quantifier $\big(\exists y/\{x\}\big)$ by storing the value of the hidden variable x in the variable z. Thus, the subformula $\big(\exists y/\{x\}\big)x = y$ has a certain meaning in the context of sentences like (1.7), and a different meaning in the context of sentences like (1.8), where variables other than x may have values.

The failure to properly account for the context-sensitive meanings of IF formulas has resulted in numerous errors appearing in the literature. We shall try to give an accessible and rigorous introduction to the topic.

Traditionally, logicians have been mostly interested in semantic games for which a winning strategy exists. Game theorists, in contrast, have focused more on games for which there is no winning strategy. The most common way to analyze an undetermined game is to allow the players to randomize their strategies, and then calculate the players' expected payoff.

We shall apply the same approach to undetermined IF sentences. While neither player has a winning strategy for the IF sentence (1.7), in a model with exactly two elements, the existential player is as likely to choose the correct element as not, so it seems intuitive to assign the sentence the truth value $1/2$. In a structure with n elements, the probability that the existential player will guess the correct element drops to $1/n$. We will use game-theoretic notions such as mixed strategies and equilibria to provide a solid foundation for such intuitions.

Chapter 2 contains a short primer on game theory that includes all the material necessary to understand the remainder of the book. Chapter 3 presents first-order logic from the game-theoretic perspective. We prove the standard logical equivalences using only the game-theoretic framework, and explore the relationship between semantic games, Skolem functions, and Tarski's classical semantics. Chapter 4 introduces the syntax and semantics of IF logic. Chapter 5 investigates the basic properties of IF logic. We prove independence-friendly analogues to each of the equivalences discussed in Chapter 3, including a prenex normal form theorem. IF logic also shares many of the nice model-theoretic properties of first-order logic. In Chapter 6, we show that IF logic has the same expressive power as existential second-order logic, and the perfect-recall fragment of IF logic has the same expressive power as first-order logic. Chapter 7 analyzes IF formulas whose semantic game is undetermined in terms of mixed strategies and equilibria. In Chapter 8 we discuss the proof that no compositional semantics for IF logic can define its sat-

isfaction relation in terms of single assignments. We also introduce a fragment of IF logic called IF modal logic.

Although it is known that IF logic cannot have a complete deduction system, there have been repeated calls for the development of some kind of proof calculus. The logical equivalences and entailments presented in Chapter 5 form the most comprehensive system to date. They are based on the work of the first author [39, 40], as well as Caicedo, Dechesne, and Janssen [9].

The IF equivalences in Chapter 5 have already proved their usefulness by simplifying the proof of the perfect recall theorem found in Chapter 6, which is due to the third author [52]. The analogue of Burgess' theorem for the perfect-recall fragment of IF logic is due to the first author. The results presented in Chapter 7, due to the third author, generalize results in [52] and extend results in [54].

Acknowledgments

The first author wishes to acknowledge the generous financial support of the Academy of Finland (grant 129208), provided in the context of the European Science Foundation project Logic for Interaction (LINT), which is part of the EUROCORES theme called "LogICCC: Modeling Intelligent Interaction." LINT is a collaborative research project that gathers logicians, computer scientists, and philosophers together in an effort to lay the grounds for a unified account of the logic of interaction.

The first and second authors would like to express their gratitude to the Centre National de la Recherche Scientifique for funding the LINT subproject Dependence and Independence in Logic (DIL) during 2008–2009, and to the Formal Philosophy Seminar at the Institute of History and Philosophy of Science and Technology (Paris 1/CNRS/ENS).

The second author is also grateful for the generous support of the Academy of Finland (grant 1127088).

The authors extend their gratitude to Fausto Barbero, Lauri Hella, Jaakko Hintikka, Antti Kuusisto, and Jonni Virtema for their many helpful comments and suggestions. They also wish to thank everyone who has helped sharpen their thinking about IF logic, including Samson Abramsky, Johan van Benthem, Dietmar Berwanger, Serge Bozon, Julian Bradfield, Xavier Caicedo, Francien Dechesne, Peter van Emde Boas, Thomas Forster, Pietro Galliani, Wilfrid Hodges, Tapani Hyttinen, Theo Janssen, Juha Kontinen, Ondrej Majer, Don Monk, Jan Mycielski,

Anil Nerode, Shahid Rahman, Greg Restall, François Rivenc, Philippe de Rouilhan, B. Sidney Smith, Tero Tulenheimo, Jouko Väänänen, Dag Westerståhl, and Fan Yang.

2
Game theory

According to *A Course in Game Theory*, "a game is a description of strategic interaction that includes the constraints on the actions that the players *can* take and the players' interests, but does not specify the actions that the players *do* take" [45, p. 2]. Classical game theory makes a distinction between *strategic* and *extensive* games. In a strategic game each player moves only once, and all the players move simultaneously. Strategic games model situations in which each player must decide his or her course of action once and for all, without being informed of the decisions of the other players. In an extensive game, the players take turns making their moves one after the other. Hence a player may consider what has already happened during the course of the game when deciding how to move.

We will use both strategic and extensive games in this book, but we consider extensive games first because how to determine whether a first-order sentence is true or false in a given structure can be nicely modeled by an extensive game. It is not necessary to finish the present chapter before proceeding. After reading the section on extensive games, you may skip ahead to Chapter 3. The material on strategic games will not be needed until Chapter 7.

2.1 Extensive games

In an extensive game, the players may or may not be fully aware of the moves made by themselves or their opponents leading up to the current position. When a player knows everything that has happened in the game up till now, we say that he or she has *perfect information*. In the present section we focus on extensive games in which the players always

have perfect information, drawing heavily on the framework found in Osborne and Rubinstein's classic textbook [45].

2.1.1 Extensive games with perfect information

Definition 2.1 An *extensive game form with perfect information* has the following components:

- N, a set of *players*.
- H, a set of finite sequences called *histories* or *plays*.
  - If $(a_1, \ldots, a_\ell) \in H$ and $(a_1, \ldots, a_n) \in H$, then for all $\ell < m < n$ we must have $(a_1, \ldots, a_m) \in H$. We call $(a_1, \ldots, a_\ell)$ an *initial segment* and $(a_1, \ldots, a_n)$ an *extension* of $(a_1, \ldots, a_m)$.
  - A sequence $(a_1, \ldots, a_m) \in H$ is called an *initial history* (or *minimal play*) if it has no initial segments in H, and a *terminal history* (or *maximal play*) if it has no extensions in H. We require every history to be either terminal or an initial segment of a terminal history. The set of terminal histories is denoted Z.
- $P\colon (H - Z) \to N$, the *player function*, which assigns a player $p \in N$ to each nonterminal history.
  - We imagine that the transition from a nonterminal history $h = (a_1, \ldots, a_m)$ to one of its successors $h^\frown a = (a_1, \ldots, a_m, a)$ in H is caused by an *action*. We will identify actions with the final member of the successor.
  - The player function indicates whose turn it is to move. For every nonterminal history $h = (a_1, \ldots, a_m)$, the player $P(h)$ chooses an action a' from the set

$$A(h) = \{\, a : (a_1, \ldots, a_m, a) \in H \,\},$$

    and play proceeds from $h' = (a_1, \ldots, a_m, a')$.

An *extensive game with perfect information* has the above components, plus:

- $u_p\colon Z \to \mathbb{R}$, a *utility function* (also called a *payoff function*) for each player $p \in N$. ⊣

Our definition differs from [45, Definition 89.1] in three respects. First, we do not require initial histories to be empty. Second, we only consider games that end after a finite number of moves. Third, we use utility functions to encode the players' preferences rather than working with

preference relations directly. We assume that players always prefer to receive higher payoffs.

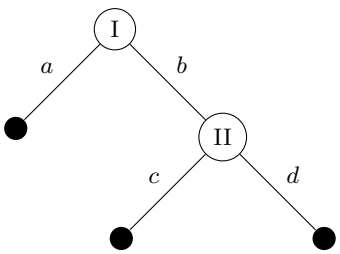

Figure 2.1 An extensive game form with perfect information

When drawing extensive game forms, we label decision points with the active player, and edges with actions. Filled-in nodes represent terminal histories. Figure 2.1 shows the extensive form of a simple two-player game. First, player I chooses between two actions a and b. If she chooses a the game ends. If she chooses b, player II chooses between actions c and d. To obtain an extensive game with perfect information, it suffices to label the terminal nodes with payoffs as shown in Figure 2.2.

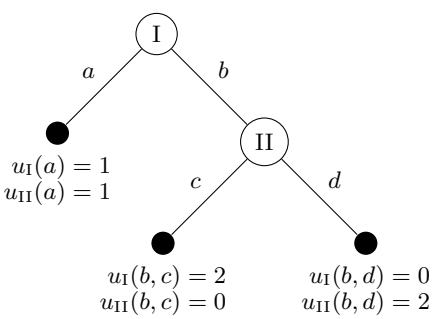

Figure 2.2 An extensive game with perfect information

Notice that the extensive game form depicted in Figure 2.1 has a tree-like structure. A *forest* is a partially ordered set $\mathbb{P} = (P; <)$ such that for all $x \in P$, the set $\{ y \in P : y < x \}$ is well ordered. The *height* of x is just the order type of $\{ y \in P : y < x \}$. A minimal element of a forest has height 0 and is called a *root*; a maximal element is called a *leaf*. The *height* of an entire forest is the least ordinal greater than the height

of every element in the forest. A *branch* is a maximal linearly ordered subset of a forest. A forest with a single root is called a *tree*.

For any two histories h and h' of an extensive game form, let $h < h'$ if and only if h is an initial segment of h'. In the game-theoretic literature, it is traditional to draw extensive game forms so that initial histories are at the top, and play proceeds down the branches. An extensive game form has *finite horizon* if the height of its set of histories is finite. All of the games discussed in this book have finite horizon.

Definition 2.2 A two-player extensive game is *strictly competitive* if the players have no incentive to cooperate, that is, if for all $h, h' \in Z$,

$$u_{\mathrm{I}}(h) \geq u_{\mathrm{I}}(h') \quad \text{iff} \quad u_{\mathrm{II}}(h') \geq u_{\mathrm{II}}(h).$$

A *constant-sum game* is one in which the sum of the players' payoffs is constant, i.e., there exists a $c \in \mathbb{R}$ such that for every terminal history h we have $u_{\mathrm{I}}(h) + u_{\mathrm{II}}(h) = c$. When $c = 0$ the game is called *zero sum*. ⊣

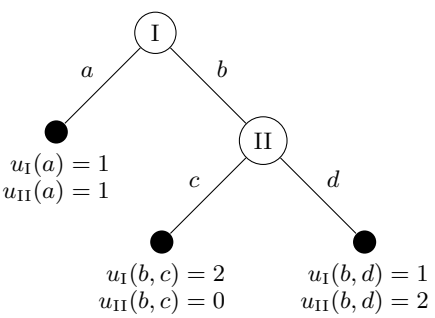

Figure 2.3 A strictly competitive game

In a constant-sum game, any gain for one player is balanced by an offsetting loss for the other. Thus the interests of the players are diametrically opposed. Every constant-sum game is strictly competitive, but not vice versa. For example, the game depicted in Figure 2.3 is strictly competitive, but not constant sum. In a zero-sum game $u_{\mathrm{II}}(h) = -u_{\mathrm{I}}(h)$ for every terminal history h.

Definition 2.3 If the only possible payoffs are 1 and 0, we say that player p *wins* a terminal history h if $u_p(h) = 1$, and *loses* if $u_p(h) = 0$. An extensive game is *win-lose* if exactly one player wins each terminal history, in which case we can replace the players' utility functions with