1
Introduction

In our everyday lives, negotiation is ubiquitous. At work, we bargain with clients about the terms of a contract; we bargain with our boss about a pay rise when the contract is signed. At home, we bargain with our partners about who will tidy the house; we bargain with our children about how many stories they can read before bed. And politicians, of course, routinely bargain in situations that have life or death consequences. The purpose of negotiation is to reach an *agreement*, and in particular, *agreement in the presence of conflicting goals and preferences*. If your preferences, goals, and aspirations were completely aligned with mine, then there would be no conflict, and hence there would be no need for negotiation. In this case, the best outcome for me would also be the best outcome for you, and so we could simply determine such an outcome, and then implement it. Neither of us would have any incentive other than to find such a mutually optimal best outcome – our joint problem is nothing more than an optimisation problem. Unfortunately, of course, the real world is not like that. More often than not, people have very different goals and preferences, and when this occurs, some method is required to find an outcome that will be acceptable to all concerned. Negotiation provides such a mechanism.

While negotiation is central to our lives, that does not imply that we are good at it – still less that we find it a pleasurable activity. When engaged in complex negotiations, people become tired, confused, and emotional, making naive, inconsistent, and rash decisions. They fall prey to personal prejudices, find it hard to weigh up complex alternatives dispassionately, and are subject to a range of misapprehensions and fallacies (Ariely, 2008). Moreover, in many cultures across the world, negotiation is felt to be a somewhat distasteful, avaricious activity. As a consequence, many people find that having to negotiate is profoundly stressful, and it is therefore no surprise that in such situations, they don't negotiate as effectively as they might in principle have been able to.

This book is about *automated* negotiation. That is, we are ultimately inter-

ested in implementing computer programs that can *negotiate on our behalf*. Because the computer programs engaged in automated negotiation are assumed to be acting on behalf of human owners or principals, we call these programs *software agents*, or (more usually) just *agents*. For reasons that should be clear from the discussion above, there are many reasons why such automated negotiators are in principle desirable; we discuss the potential advantages (and disadvantages) of automated negotiation in more detail below.

The overall aim of this book is thus to describe the challenges that arise when we want to implement software agents that can negotiate, to describe the main approaches to resolving these challenges, and to give a feel for the potential applications of future automated negotiation technologies.

1.1 The structure of negotiation

If we want to build computer programs that can negotiate effectively, then it is natural to consider first how people negotiate. Negotiation has, after all, been studied widely in areas as diverse as management science and economics, psychology, and international relations – what can research in these disciplines tell us about how to construct negotiating computer programs? Lopes *et al.* (2008) discuss how negotiation has been studied in these disciplines, and on the basis of this work, suggest that it can be understood as consisting of eleven overall activities, as follows:

1. *Social conflict*. The negotiation process begins with the recognition that there is some *conflict* that needs resolution, and by determining the extent and scope of this conflict, including the issues in contention. Note that we do not mean here the term conflict in the sense of war. We simply mean that there are two or more individuals who have differing preferences over some matter, and that they have some common purpose in resolving this situation. The conflict could be anything from two children trying to decide how to divide a cake, to nations trying to decide who should have sovereignty over a piece of disputed territory.
2. *Negotiation participants*. At this stage, the stakeholders in negotiation are identified – those that have an interest in the outcome, and that will participate in the negotiation process.
3. *Gathering and structuring of private information*. The planned participants of negotiation will at this stage evaluate and fix their own priorities and preferences with respect to the issues at stake. For example, if you are aiming to purchase a second-hand car, this stage might involve gathering information about the reliability of various makes of cars, researching

purchase prices for various models of cars in various conditions, and then deciding which features are important to you, forming priorities among them, and so on.

4. *Opponent analysis*. Here, the planned participants consider the other negotiation participants. In human negotiations, this may involve "intelligence gathering" to determine how other participants rate the various issues, as well as how other negotiation participants are likely to behave during negotiation.
5. *Protocol selection*. At this stage, the actual "rules of encounter" (cf. Rosenschein and Zlotkin, 1994) are fixed for the negotiation. These rules will determine how negotiation will proceed (e.g., the order in which participants are permitted to act), the space of possible proposals, the rules determining how such proposals may be made, and how agreement or failure is determined.
6. *Exchange of offers and feedback information*. This step constitutes the heart of the actual negotiation process itself. Typically, it consists of an iterated exchange of offers and counter offers, until either agreement is reached or negotiation breaks down. The behaviour of negotiation participants during this stage – the process of making offers and counter offers – is governed by their *negotiation strategy*, of which we will say much more in the remainder of this book.
7. *Argumentation (justifications, promises, etc.)*. Everyday human negotiation is often accompanied by *arguments* of various kinds. The role of such arguments may be to present *justifications* for the proposals being made ("I can't look after the children tomorrow because I'm working late", "I can't offer any more for the car because…"). In human negotiations such arguments may also include threats or promises.
8. *Learning*. As negotiation proceeds, information is usually revealed to participants – in particular, information about other participants, which may replace or supplement information that was previously believed about these participants. For example, we may obtain new insights into how other participants value the various issues at stake. Learning in this context means the process of adding to and updating the body of beliefs we have about others.
9. *Dynamic strategy selection*. As new information comes to light while negotiation proceeds, we may decide to alter our originally chosen negotiation strategy. Of course, if we think of a strategy as dictating how we should negotiate for every conceivable situation, then we aren't really changing our strategy as new information comes to light; but in practice, most people will not define a strategy for every possible circumstance, but simply plan

how to act on the basis of the beliefs they currently have, and then reconsider their plan if these beliefs are demonstrated to be false.

10. *Impasse resolution*. An impasse occurs when negotiation breaks down, and no further progress can be made towards a resolution through the agreed negotiation procedure. At this stage, alternative action may be required to enable an agreement to be reached (e.g., bringing in a third-party mediator).
11. *Renegotiation*. This step allows for refinements to be made to an agreed deal, for example if it is seen that a solution is in some sense inefficient. The key measure of efficiency in this context is called *Pareto optimality*, and we will see that this plays an important role in the negotiation protocols considered in this book.

Of course, not all instances of negotiation will encompass all these stages, some of which will be trivial or redundant for certain domains. In this book, we will focus on a strict subset of these stages: the main issues of interest are described in the following section.

1.2 Parameters of automated negotiation

In this book, we consider a negotiation setting to be defined by three components: the *negotiation set*, the *negotiation protocol*, and the *strategies* that may be used by negotiation participants.

The negotiation set. The negotiation set represents the space of possible proposals, or "deals" that agents can make. Defining the negotiation set is one of the key requirements for negotiation, automated or otherwise. A deal will be characterised by one or more attributes, taking values from some domain (e.g., when purchasing a car the attributes might be price, colour, delivery time, length of guarantee, and so on). We typically regard domains that are characterised by a single issue as being simpler than domains characterised by multiple issues.

An example of a single-issue negotiation scenario might be where a buyer and a seller are negotiating on the price of a particular item for sale. In such a scenario, the preferences of the agents are symmetric, in that a deal which is more preferred from the point of view of the seller is guaranteed to be less preferred from the point of view of the buyer, and vice versa. Such symmetric scenarios are relatively simple for most people to understand, because it is always obvious what represents a concession: in order for the seller to concede, he must lower the price of his proposal, while for the buyer to concede, he must raise the price of his proposal. In multiple-issue negotiation scenarios,

agents negotiate over not just the value of a single attribute, but also the values of multiple attributes, which may be interrelated. In multiple-issue negotiations, it may be hard for people to compare different deals dispassionately, and professional salesmen often exploit this fact during negotiation by making "concessions" that are in fact no such thing.

Another complication of multiple-issue negotiation is that multiple attributes lead to an exponential growth in the space of possible deals. Consider a (somewhat artificial) domain in which agents are negotiating over the value of n Boolean variables. A deal in such a setting consists of an assignment of either true or false to each variable. There are 2^n possible deals in such a domain. This means that it would not be feasible for an agent to consider every possible deal explicitly in domains of moderate size, and hence finding an *optimal* deal for the agent from the set of possibilities may be computationally hard. Most negotiation domains are, of course, much more complex than this. In real-world negotiation settings – such as labour disputes or negotiations between nations – the issues of negotiation are very complex indeed: laws, regulations, and so on. In such cases, it can be difficult even to define an appropriate mathematical model of the domain.

Worse still, the negotiation participants may have difficulty reaching agreement on what the attributes under negotiation actually are. To consider a real-world example, in the early part of the 21st century, delicate negotiations were under way in Northern Ireland between the UK government and various paramilitary organisations, with respect to the future of the beautiful but historically troubled region. One of the key issues in these negotiations was that of decommissioning weapons that had previously been used by paramilitary forces. It became clear at various points that the relevant parties had quite different views on whether the decommissioning of paramilitary weapons should be up for negotiation or not. Fortunately, in the end, agreement was reached – a happy example of successful negotiation.

Thus, the negotiation set effectively circumscribes the domain of negotiation. It defines both the issues that are up for negotiation, and the values that these issues can take. Chapter 4 is devoted to the subject of how negotiation domains can be formalised in a way that is amenable to automated negotiation. While not wishing to pre-empt that chapter, at this stage we should say a little about the kinds of negotiation domains we will and will not be considering in this book. In particular, we will *not* be considering the problem of *negotiating over beliefs*. Such negotiation occurs when two or more individuals have conflicting beliefs, and need to come to a consistent interpretation of these conflicting belief sets. Such a situation occurs, for example, in a court of law, when various parties put forward positions that are, in a logical sense,

strictly contradictory – the task of the jury and/or judge is to arrive at a consistent interpretation of these conflicting positions. Although such "negotiation" falls outside the scope of this book, it does fall within the scope of artificial intelligence research, where it is the purview of the *argumentation* research area. See Besnard and Hunter (2008), Rahwan and Simari (2009) for excellent introductions to contemporary research on argumentation. Although it falls outside the scope of this book, we briefly comment on argumentation research in Chapter 13. The kinds of domains we will focus on tend to relate to the *sharing of scarce resources* or *sharing the benefits of cooperation*. We will see this in more detail in Chapter 4.

The negotiation protocol. The negotiation protocol constitutes the set of rules that define how negotiation will proceed. In particular, it defines the proposals that agents are allowed to make, as a function of prior negotiation history. Although negotiation protocols differ in some respects, the protocols we consider in this book closely resemble the negotiation we are familiar with from everyday life and work. Thus, negotiation typically proceeds over a series of rounds, with one or more proposals being made at each round. As we will see, in many cases, an analysis of the negotiation scenario by participants enables them to reach an agreement immediately.

The particular protocol in use will often involve rules that impose constraints on the deals that can be made. For example, there might be a rule stating that participants are not permitted to repeat a previous proposal. Such a rule prevents "filibustering", where one participant tries to delay the negotiation process. In domains where the negotiation set is finite, this rule can ensure that if negotiation proceeds long enough, then eventually all possible deals are considered.

The negotiation protocol will typically include a rule that determines when a deal has been struck, and what this *agreement deal* is, and will also include a rule that determines the situations under which negotiation will be deemed to have failed, and define the *conflict deal*: the outcome that will be implemented if negotiation fails.

Negotiation strategies. A negotiation strategy for a participant defines the proposals that this participant will make over time: how they will negotiate, in other words. Usually, the strategy that an agent uses is *private*: the fact that an agent is using a particular strategy is not generally visible to other negotiation participants (although most negotiation settings are "open", in the sense that the actual proposals made *are* seen by all participants). When we negotiate informally in everyday life, we may not have formulated a detailed strategy, but

as we move to more formal negotiation settings, it is likely that participants will invest more time and energy into formulating their strategy.

Other parameters. Another parameter of negotiation is the number of agents involved in the process, and the way in which these agents interact. There are three obvious possibilities:

- *One-to-one (bilateral) negotiation.* In this setting, one agent negotiates with just one other agent. A particularly simple case of one-to-one negotiation is that where the agents involved have directly contradictory preferences with respect to the possible deals. An example from everyday life would be the type of negotiation we get involved in when discussing the price of a car with a car salesman. We will see examples of such symmetric negotiation scenarios later.
- *Many-to-one negotiation.* In this setting, a single agent negotiates with a number of other agents. For the purposes of analysis, many-to-one negotiations can often be treated as a number of concurrent one-to-one negotiations. They can also be treated as auctions. Although auctions are somewhat beyond the scope of this book, we do provide a brief introduction later.
- *Many-to-many negotiation.* In this setting, many agents negotiate with many other agents simultaneously. In the worst case, where there are n agents involved in negotiation in total, this means there can be up to $n(n-1)/2$ negotiation "threads" in progress at any given time. These settings are often analysed using market theories, and are somewhat outside the scope of the present book.

1.3 A strategic approach

For most of this book, we adopt the standpoint that negotiation may be fruitfully understood as being a *strategic* activity. That is, we assume that when participants engage in negotiation, these participants act as best they can to realise their preferences, assuming that their counterparts in turn will act as best they can to realise their preferences. *Game theory* is the mathematical theory of strategic decision making (Maschler *et al.*, 2013). If we take seriously the view that negotiation is strategic, then it seems that game theory might be an appropriate analytical tool with which to understand how strategic negotiators can and should act, and might also be useful in both the design of negotiating agents and protocols for automated negotiation.

Game theory was originally developed to study recreational games such as

chess and poker, and much terminology in the field is derived from such games (including, of course, the term "game theory" itself). However, game theory is *not* just concerned with recreational games. Game theory is concerned with all settings in which self-interested agents interact with each other in pursuit of their own preferences, taking into account each other's strategic decisions. A "game" in the sense of game theory is an abstract mathematical model of a strategic decision-making setting. It is abstract in the sense that it contains only information that is germane to the decisions that players must make. Typically, a game might specify:

- who the players in the game are (i.e., the agents that are interacting);
- what choices are available to the players;
- what the beliefs of the players are (the information they have about the game and about other players);
- what the outcomes of the game will be, as a function of the choices made by players; and
- what the preferences of the players are with respect to the possible outcomes.

A key concern in game theory is to identify what the outcomes of games will be, under the assumption that players act rationally in pursuit of their preferences. To this end, game theory provides *solution concepts*, which typically attempt to characterise some notion of rational choice in a game. A solution concept singles out some outcomes of the game as being those that would be selected if the players used the corresponding notion of rational choice. Probably the best known such solution concept is *Nash equilibrium*, and indeed we will be hearing a lot about Nash equilibrium in this book. A fundamental problem in game theory is that solution concepts often fail to identify a *unique* outcome of the game: it may be, for example, that there is no rational outcome of the game according to a particular solution concept (in which case, what use is the solution concept with respect to a player's decision problem?), or it may be that there are *multiple* outcomes of the game according to the solution concept (in which case, how should individual agents make independent choices so as to collectively select one of these solutions?). Historically, much game theory research has been concerned with formulating and refining solution concepts, in particular with the goal of resolving the problem of having multiple solutions to a game.

We should be clear about the role that we expect game theory to play in our work. Game theory is usually interpreted in one of two ways: as a *descriptive* theory or as a *normative* theory. Under the descriptive interpretation, game-theoretic analysis is interpreted as making predictions about how people will act

in strategic decision-making situations. In contrast, the normative view interprets game theory as providing advice about how players should act in strategic settings – essentially, it tells players how they ought to act.

Both of these interpretations have been subject to criticism. With respect to the descriptive interpretation, there seems at first sight to be a wealth of evidence suggesting that game theory does not work well in predicting how people will act in strategic settings. Binmore (2007) considered this question in great detail. He concluded that there *are* situations in which game theory can work as a predictive theory for human decision making, but that for it to do so, the scenario to which game theory is being applied must satisfy the following properties:

- The scenario should be sufficiently simple that the decision makers can understand it. That is, they need to be able to understand easily the choices available to themselves and others, the incentives of others (i.e., the preferences they are trying to realise), and the consequences of various choices.
- The incentives driving the decision makers should be sufficiently large that they really do affect the decisions made by participants.
- The participants must be able to repeat the interaction, observe the outcomes, and learn from them: if a player interacts just once, then they don't have sufficient opportunity to learn. Over time, Binmore (2007) suggests, the choices of players will converge towards game-theoretic solution concepts.

Some further discussion of these points is provided by Wooldridge (2012).

Our present work, which aims to use game theory to inform the design of software negotiators, clearly has more of a focus on the normative interpretation of game theory than on the descriptive interpretation. Here, we have one advantage over those that aim to reconcile game theory with observed human behaviour. If we are going to build software agents that will interact with other software agents, then we can *design* these agents to be rational decision makers. For example, we don't have to worry about negotiation participants getting bored, distracted, or tired, because they are computer programs. We can design our programs to be systematic, rigorous, and strictly rational in their decision making (subject to the limitations imposed by available processor power and memory space). These are not assumptions that can always be made legitimately about people when they negotiate. Of course, when our negotiating software agents must negotiate with people, then that changes the flavour of the interaction somewhat – Chapter 10 is devoted to the issues that arise when software agents must negotiate with people.

1.4 Desiderata for automated negotiation

Rosenschein and Zlotkin (1994, pp. 20–22) suggest some criteria that we should require of any framework for automated negotiation. The criteria they propose (with a few additional points) are as follows:

- *Economic efficiency*. We have already alluded to this property above, in the context of possibly renegotiating to improve the negotiation outcome. We say that a negotiation outcome is *in*efficient if there is some participant i that can look at that outcome and point out that there is another outcome that would make i better off without making anybody else worse off. If an outcome is not inefficient in this sense, then we say it is economically efficient. Economic efficiency in this sense – technically, *Pareto optimality* – seems a natural and easily motivated property. However, it is important to note that while an outcome being Pareto efficient is a *necessary* condition for that outcome to be considered a reasonable outcome of negotiation, it is not a *sufficient* condition. It is easy to construct perverse Pareto-efficient outcomes. For example, suppose two participants are negotiating over how to divide a dollar. Then the outcome that gives the whole dollar to one player is efficient in the sense we describe above, clearly isn't very desirable from the point of view of the other player. Thus, the concept of Pareto efficiency is not concerned with *fairness*.
- *Individual rationality*. Our usage of this term derives from game theory, and we will define it formally later. Informally, the idea is that players have an active incentive to participate in the negotiation process: no player could end up being *worse off* as a consequence of participating than they were before negotiation began.
- *Simplicity*. We mean this in two senses. First, we mean that given a negotiation protocol, it should be "obvious" to players what their best strategy is. In computational terms, "obvious" means that it should be possible to compute their best negotiation strategies in polynomial time. Second, we mean that once participants have determined their optimal strategy, it should be easy for them to *follow* this strategy. In computational terms, this means the strategy itself should be computable in polynomial time.
- *Distribution*. A negotiation protocol is distributed if it can be enacted without the need for a third party being involved. Thus, a distributed protocol is one that can be enacted by the participants alone. The rationale for requiring distribution is primarily that of engineering design: it avoids introducing a single point of failure, or bottleneck in the protocol. However, developing