Joint Practical Deliberation

by

Brendan de Kenessey

B.A., Yale University

Submitted to the Department of Linguistics and Philosophy
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Signature of Author..

Signature redacted

Brendan de Kenessey
Department of Linguistics and Philosophy
May 31, 2017

Signature redacted

Certified by....................................

Kieran Setiya
Professor of Philosophy
Thesis Supervisor

Signature redacted

Accepted by....................................

Roger White
Professor of Philosophy
Chair of the Committee on Graduate Students

Joint Practical Deliberation
by
Brendan de Kenessey

Submitted to the Department of Linguistics and Philosophy
on May 31, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Philosophy

Abstract

Joint practical deliberation is the activity of deciding together what to do. In this dissertation, I argue that several speech acts that we can use to alter our moral obligations – promises, offers, requests, demands, commands, and agreements – are moves within joint practical deliberation.

The dissertation begins by investigating joint practical deliberation. The resulting account implies that joint deliberation is more flexible than we usually recognize, in two ways. First, we can make joint decisions not only about what *we* will do together, but also about what *you* or *I* will do alone. Second, we can deliberate by means of two distinct methods: *propose-and-ratify*, in which a proposed joint decision must be explicitly accepted to come into effect, and *propose-and-challenge*, in which a proposed joint decision comes into force unless it is explicitly challenged.

Varying these parameters generates a botany of different kinds of proposals we can make within joint deliberation. When we look at these proposals more closely, we make a surprising discovery: for each kind of proposal we can make in joint practical deliberation, there is an everyday speech act with the very same properties. A certain kind of proposal to make a joint decision regarding one's own actions has the same normative effects, under the same conditions, as a promise. One kind of proposal to make a joint decision regarding one's addressee's actions has all the essential features of a command; another kind of deliberative proposal – with the same content but a different method of evaluation – looks exactly like a request. And so on.

These similarities are too systematic to be coincidental. The only explanation, I argue, is that these ordinary speech acts are identical to their doppelgängers within joint practical deliberation. *Promises* and *offers* are proposals to make joint decisions about what *I* will do. *Commands, demands,* and *requests* are proposals to make joint decisions about what *you* will do. And *agreements* are joint decisions about what *we* will do. Call this the *deliberative theory* of these speech acts.

Considering each speech act in turn, I defend the deliberative theory by arguing that it provides a uniquely powerful explanation of its targets' social and moral significance. Once we see how naturally these speech acts fall out of our practice of joint deliberation, theories that treat them as *sui generis* – as many moral philosophers now do – will come to seem redundant and nonexplanatory. Conversely, thinking of promises, offers, commands, demands, requests, and agreements as moves within joint practical deliberation allows us to give an elegant and generative theory of these phenomena that have confounded moral philosophers for so long.

Thesis Supervisor: Kieran Setiya
Title: Professor of Philosophy

Table of Contents

Acknowledgements

Introduction

A few words can make a big moral difference. If I say, "I'll come to your talk, I promise," this transforms my later absence from a mild disappointment into a breach of trust. When the commander yells, "Retreat!" the choice to charge ahead is no longer merely imprudent: it is now insubordination. If a recovering addict asks you to be her sponsor, then in simply saying "Yes," you undertake a life-altering responsibility.

These cases illustrate a puzzling fact: some obligations are such that we can create them seemingly at will. This is puzzling because so many obligations are *not* this way: for example, nothing anyone says can alter my obligation to not inflict pointless suffering. But when I make a promise, offer, or demand, accept a request or agreement, or give a command, I seem to change what morality requires simply by declaring it to be so.[1] How could this be?

In this dissertation I will defend a novel answer to this question: these speech acts change our obligations by changing what we have *jointly decided* to do.[2] When you and I decide that we will get lunch tomorrow, we each become obligated to the other to show up for lunch. This is not so puzzling: it is because we are both obligated to follow through on the plans we make with others. What we have changed by declaration is not directly what morality requires, but instead what our joint plan is – which in turn morality requires us to follow.

I think that *all* of the above examples can be explained in this way. To see this, we have to understand just how flexible and pervasive joint decision-making is.

---

[1] Conspicuously absent from this list is *consent*, another speech act with the power to change morality by declaration. Consent has the power to make otherwise wrongful actions permissible, as when one consents to surgery or sex. Like the other speech acts covered in this dissertation, I believe that consent is best understood as a move within joint practical deliberation. In particular, I think that consent is a certain kind of proposal to *retract* a joint decision that is already in force. To defend this hypothesis, however, I would need to defend its implications for the nature of the rights that consent waives. Given the centrality of rights in moral philosophy – and the massive literature on the subject – doing justice to this topic would probably require another dissertation. Sadly, consent will have to wait.

[2] Following convention, I will use the term 'speech act' to refer to the kind of communicative action of which promises, assertions, questions, commands, etc. are paradigm cases. But I do not assume that such acts can only be performed using speech: I might make a promise with a significant nod, or ask a question by raising an eyebrow. Thanks to Caspar Hare and Matt Mandelkern for suggesting I clarify this point.

The first step is to observe that we can make joint decisions not just about what *we* will do, but also about what *you* will do or what *I* will do. If we are building a fire, we might jointly decide that you will gather the kindling, leaving it quite open what I will do. This could happen in at least two ways. You could offer to gather the kindling, and I could accept your offer. Or, I could ask you to gather kindling, and you could accept my request. The result of both these exchanges is a joint decision to the effect that you will gather the kindling.

The second step is to distinguish two methods of joint decision-making (see Ch. 1, §3.1). Offers and requests employ what I call the *propose-and-ratify* method: the speaker proposes a joint decision, and then the addressee(s) must explicitly accept the proposal for it to come into force. But we can also use what I call the *propose-and-challenge* method, in which a proposed joint decision comes into force by default unless it is explicitly challenged. This, I suggest, is how promises, demands, and commands work. Promises propose joint decisions regarding one's own actions by means of propose-and-challenge: "I'll gather the kindling." Commands and demands propose joint decisions regarding another's actions using propose-and-challenge: "Go get the kindling!"

My thesis is that promises, offers, agreements, requests, commands, and demands are all moves within *joint practical deliberation*, the activity of deciding together what to do. *Promises* and *offers* are proposals to make joint decisions about what *I* will do. *Commands, demands,* and *requests* are all proposals to make joint decisions about what *you* will do. *Agreements* are joint decisions regarding what *we* will do. Call this the *deliberative theory* of these speech acts. In the chapters to come, I will defend the deliberative theory by showing how it offers a uniquely powerful explanation of these speech acts' social and moral significance. Once we see how naturally these speech acts fall out of our everyday practice of joint deliberation, theories that treat them as *sui generis* – as many moral philosophers now do – will come to seem redundant and nonexplanatory.

The deliberative theory makes claims of the form "X is Y": e.g., "promises are proposals to make joint decisions about one's own actions." The theory aims to illuminate various phenomena we recognize in ordinary language (promises, commands, etc.) by identifying them with elements in a more abstract theory (proposals in joint practical deliberation). In this sense,

the deliberative theory offers an *analysis* of these speech acts: it tells us *what they are*. On what basis might we accept such an analysis?

Compare an analysis we *do* accept: the thesis that temperature is mean kinetic energy. (I am told that the truth is a bit more complex; but that should not matter for our purposes). Suppose you are an eighteenth-century physicist developing a theory of molecular motion. As you perform your calculations, you come to a surprising realization: your theory predicts that the *mean kinetic energy* of the molecules that compose a substance will be perfectly correlated with that substance's *temperature*. Just as increasing a gas's pressure will heat the gas up, so too it will increase the gas's mean kinetic energy; just as decreasing the temperature of water makes it freeze, so too will decreasing its mean kinetic energy. What could explain these observations? The best explanation seems to be that temperature and mean kinetic energy are the same property. After all, you might reason, it is unlikely that there are two separate properties, temperature and mean kinetic energy, which coincidentally happen to have the very same characteristics. And by accepting their identity, you gain a great deal of explanatory power. You can now explain temperature's various quirks and qualities by showing how they can be derived from the laws of molecular motion (cf. Block & Stalnaker 1999: 23-24). Where before it was an isolated phenomenon, practically essential but theoretically opaque, temperature now becomes an integrated part of our wider understanding of how particles interact.

My argument for the deliberative theory takes the same form. I begin by providing an independently motivated theory of joint practical deliberation. (In our analogy, this plays the role of the theory of molecular motion). My account of this activity, presented in Chapter 1, is based on two main ideas. First, joint practical deliberation has the same structure as *individual* practical deliberation, the activity of deciding what to do alone. Second, joint practical deliberation cannot require its participants to give up their individual rationality or autonomy. Teasing out the consequences of these two constraints, we can derive a detailed account of joint practical deliberation's nature and norms.

Varying the parameters of this theory, we can generate a botany of different kinds of proposals that can be made within joint deliberation. We can then use the theory to predict the normative and communicative properties each proposal will have. When we do so, we make a surprising discovery: for each kind of proposal we can make in joint practical deliberation, there is an ordinary, everyday speech act with the very same properties. A certain kind of proposal to make a joint decision regarding one's own actions has the same normative effects, under the same conditions, as a promise. One kind of proposal to make a joint decision regarding the addressee's actions has all the essential features of a command; another kind of proposal – with the same content but a different method of evaluation – looks exactly like a request. And so on.

These similarities are too systematic to be coincidental. The best explanation, I argue, is that these ordinary speech acts are identical to their doppelgängers within joint practical deliberation. After all, it is unlikely that there are two separate sets of speech acts that coincidentally happen to have the very same characteristics: promises, requests, etc. on the one hand, and proposals in joint deliberation on the other. And by accepting this analysis, we gain a great deal of explanatory power. We can now explain each speech act's distinctive quirks and qualities by showing how they can be derived from the laws of joint practical deliberation. Where before they were isolated phenomena, practically essential but theoretically opaque, promises, offers, requests, agreements, demands, and commands now become an integrated part of our wider understanding of how people interact.

The rest of the dissertation applies this explanatory strategy to each speech act in detail.

We begin, in Chapter 2, with promises. Out of the speech acts that the deliberative theory targets, promises have received the most philosophical attention by far. Thus I develop the deliberative theory of promises in the most detail, showing how it predicts and explains the well-studied nuances of this speech act. Why is the obligation to keep a promise owed to the promisee in particular? Because the joint decision in which this obligation is based was as much the promisee's as it was the promisor's. How is it possible for the promisee to release the promisor from her obligation? She can do so by proposing to retract the joint decision the promise brought

into force. Why are promises induced by coercion or deception not binding? Because if you coerce or deceive me into making a promise, you cannot be deliberating with me in good faith, and this renders null and void any 'joint' decisions we reach.

Chapter 3 extends the deliberative theory to offers, agreements, and requests − speech acts that have received far less philosophical discussion. I point out that these speech acts are all deeply similar to promises, sharing promises' characteristic pattern of normative force and validity conditions. This is just as the deliberative theory would predict: since promises, offers, agreements, and requests all have the effect of bringing a joint decision into force, we should expect their similarities to be more significant than their differences. Other theories have a harder time explaining the unity of these speech acts. The only way they can do so, it seems, is by claiming that offers, agreements, and requests are special kinds of promises. But, I argue, this reductive view fails to capture the features that make offers, agreements, and requests distinctive.

In Chapter 4, we turn to demands and commands. I argue that these two terms of ordinary English pick out the same speech act employed in different situations. Both commands and demands are proposals to make joint decisions regarding the addressee's actions using the propose-and-challenge method. We call such proposals *demands* when they propose actions that their addressees were already obligated to do: as when you come across me stomping on a child's sand castle and demand, "Stop that at once!" And we call these proposals *commands* (or, equivalently, *orders*) when they propose actions that their addressees would otherwise not have been obligated to do: as when a sergeant commands her soldiers, "Drop and give me fifty!" (Push-ups, that is). In a slogan: commands create new obligations, while demands hold people to the obligations they already have. I argue that the deliberative theory explains the common force of these speech acts while predicting the differences between them.

Beginning with demands, I show how the deliberative theory helps to capture our intuitive judgments about what we can legitimately demand of one another. There are some actions I ought to do, but nobody can *demand* that I do: brush my teeth, eat my vegetables, cultivate my painting hobby. There are other actions that particular people seem to have special

standing to demand of me: my mother can demand that I call her on her birthday, my students can demand that I grade their papers fairly. Finally, there are actions it seems that *everyone* can demand of me: that I not stomp on children's sand castles, or dump poison into the ocean. Why can people demand some actions of me but not others? It has to do, I suggest, with what reasons I *share* with others in the sense relevant to joint practical deliberation. By looking at the shared reasons we take up in joint deliberation, I argue, we can predict when others will have the standing to make demands of us, and when they should mind their own business.

The deliberative theory of commands is surprising. It is odd to describe commands as proposing *joint* decisions, since commands involve a steep asymmetry in power between speaker and addressee. I argue that this asymmetry arises because commands employ the propose-and-challenge method. Propose-and-challenge gives the speaker disproportionate power: her proposals come into force by default unless someone raises a convincing objection to them. But sometimes we have good reason to give someone this kind of power. If we are unloading a moving van, for example, things will go more smoothly if we make one person the 'leader', giving her exclusive power to direct traffic by propose-and-challenge. By asking when we ought to give someone this deliberative power, we can derive a predictive account of what it takes to have legitimate authority.

The speech acts this dissertation investigates are of more than just theoretical interest. They play a critical role in our social lives and moral thought. Our daily schedules are structured around the commitments we make using promises, offers, agreements, and requests. You might spend the morning reading a draft that your student requested comments on, go to the lunch meeting you agreed to attend, drive to pick up the friend whom you offered a ride to the airport, and spend the evening taking your daughter to the movies, as you promised her you would.

Commands and demands may be less salient in daily life, but are of no less significance. Many of the institutions of modern civilization – governments, universities, corporations – are built on the idea of authority to command. And in our personal lives, we pay careful attention to

the question of what our friends, family, and colleagues can demand of us. The thought that we might not live up to one of these demands has a motivating force unlike almost any other.

This dissertation is thus not just about the philosophical puzzle of how it is possible to change our obligations with words. Its true aim is to build a framework within which we can locate and illuminate these forms of speech that carry so much weight in our daily lives. If you have ever gone out of your way to fulfill a promise, been excited by an offer, hashed out the terms of an agreement, felt nervous about making a request, resented the commands of an authority, or been moved by the thought of what someone could demand of you, this dissertation is an attempt to better understand your life.

Chapter 1

Joint Practical Deliberation

Joint practical deliberation is the activity of deciding together what to do.[1] When friends discuss where to meet for coffee, when doctors debate whether to perform a risky surgery, when partners talk about whether to have a third child, they are engaged in joint practical deliberation. They are trying to arrive at a single plan of action that expresses and governs all of their wills.

My account of this activity will be based upon two guiding ideas. The first idea is that joint practical deliberation is an activity of the very same kind as *individual* practical deliberation, the activity people engage in when they deliberate alone about what to do. Like making lasagna or going for a walk, practical deliberation is the kind of activity that can be performed both together and alone. The second idea is that the norms of joint practical deliberation must be compatible with the requirements of individual rationality and autonomy. It must be possible to fully participate in the activity of joint practical deliberation without giving up one's individual commitment to acting rationally and autonomously.

Together, these two ideas suggest a strategy for investigating joint practical deliberation. The first idea suggests that joint practical deliberation will have the same structure as individual practical deliberation. So, we should begin by looking at individual practical deliberation, and asking what it would take for two or more agents to engage in a single activity with the same structure.[2] We can then leverage the second idea to fill in the details of this structure, asking what the norms of joint deliberation would have to be in order to be compatible with each of its

---

[1] A note to the reader. This chapter lays the foundation for the whole variety of arguments and applications that will occupy the remainder of the dissertation. As a result, however, it is an exercise in delayed gratification. To develop all the tools we need, the following sections must cover details the relevance of which will not become fully clear until later in the dissertation. Readers who are anxious to get straight to the applications are welcome to read the interim summary in §4 and then skip to the presentation of the deliberative theory in §6. You can then refer back to the other sections of this chapter as needed to understand later arguments.

[2] This strategy distinguishes my account from other work on joint practical deliberation (e.g., Westlund 2009; Laden 2012; Bratman 2014: Ch. 7): these philosophers put much less weight on the idea that joint deliberation is analogous in structure to individual deliberation.

14

participants' rationality and autonomy. This chapter will pursue this strategy, pulling a theory of joint practical deliberation out of the constraints imposed by these two basic premises.

First, a methodological note. You might have expected that our account of joint practical deliberation would have to begin by investigating the nature of *joint activity* in general. In other words, you might think we need to begin by answering the question: what does it take for two or more agents to do something *together*, rather than each acting separately? But, while this question is important, I don't think we need to answer it here.

To see why, suppose we are investigating some other everyday joint activity: for example, the activity of *going on a date*. The questions we are likely to ask will concern the features that distinguish dates from other kinds of joint activities. What is the difference between going to the movies *as a date* and going to the movies *as friends*? What are the common features that make disparate date activities – going to the movies, eating a meal, touring an art gallery – all count as dates? But it seems clear that we can answer these questions without saying what it is that makes a date a *joint* activity. In fact, the metaphysics of joint activity seems irrelevant to our inquiry. Instead, we should take the fact that dates are joint activities for granted, and focus on the specific structure and norms that make dating unique.

I think we should approach joint practical deliberation in the same way. Instead of worrying about what makes joint practical deliberation a *joint* activity, we should take the jointness of this activity for granted, and focus on elaborating its specific structure and norms. Proceeding in this way, we can learn a great deal about the nature of joint practical deliberation while saying almost nothing about the metaphysics of joint action. Indeed, as far as I can tell, the theory of joint practical deliberation offered in this chapter is compatible with any of the theories of joint action currently on the market (e.g., Tuomela 2007, Gilbert 2011a, and Bratman 2014).

1. The structure and elements of joint practical deliberation

Let us begin by looking at individual practical deliberation, and then ask what it would take for multiple persons to engage in a single activity with the same structure.

What is the structure of individual practical deliberation? Consider a concrete case: I am deciding where to get lunch. I am faced with a set of *options*, the potential actions I am choosing between (to go to Clover, to Mexicali, or to Darwin's).[3] My aim is to make a *decision* that narrows those options down to one (say, to go to Mexicali). However, my decision will (typically) not be arbitrary – it will be based on *reasons*. For example, the fact that Clover sources its food from local farmers might give me some reason to go there, while the fact that I'm craving a burrito might give me more reason to go to Mexicali, which makes a mean burrito.

Putting these elements together, we can characterize individual practical deliberation as the process of making a *decision* about which of one's *options* to perform on the basis of one's *reasons*. By analogy, we can conclude that joint practical deliberation is a process of making a *joint decision* about which of our *shared options* to perform on the basis of our *shared reasons*.

This schema already tells us something substantive about joint practical deliberation: we should reject the game-theoretic idea that joint deliberation is merely a negotiation between individuals who each take into account only their own preferences and beliefs. If we take the analogy between joint and individual deliberation seriously, then we should see joint deliberation as a collective undertaking where the participants start with a *shared* set of options and reasons and work together to reach a single decision that represents all of their wills (cf. Westlund 2009 and Laden 2012: 184-197).

Let us treat the italicized phrases in our schema as functional terms: 'shared options,' 'shared reasons', and 'joint decisions' each refer to *whatever it is* that plays the same role in joint practical deliberation its analogue plays in individual practical deliberation. Our question now is what can play the roles picked out by these terms. Let's consider each in turn.

---

[3] Clover, Mexicali, and Darwin's are all restaurants nearby the MIT philosophy department.

1.1. Shared options

In the individual case, a person's options are the possible actions she might perform. This suggests an idea: in parallel with the popular possible worlds framework for modeling theoretical deliberation (e.g., Stalnaker 1984), we can model practical deliberation using a set of *possible actions* (cf. Gibbard 2003: Ch. 3).[4] On this model, the content of a decision (to go to Mexicali) is a set of possible actions (the set of all actions in which I go to Mexicali), and the effect of this decision is to exclude from future consideration any possible actions outside of that set (e.g., going to C instead). A person's *options* are the result of partitioning her possible actions into mutually exclusive subsets, each of which is the potential content of a decision. Thus one of my options might be the set of possible actions in which I go to Mexicali, another the set of possible actions in which I go to Darwin's, and so on.

This model is easily extended to the case of shared options. Instead of considering possible actions one person can perform, we can consider the possible *combinations of actions* multiple persons could compatibly undertake. This will include not only paradigm joint activities, but also *any* possible combinations of the relevant people's individual actions, such as the combination in which I teach a class in Paris while you deliver a baby in Los Angeles. Let us model the shared options of multiple persons as mutually exclusive sets of possible combinations of their actions.[5] Joint decisions take one of these sets as their content (say, the set in which we go to Mexicali together), and exclude from joint consideration combinations of actions that fall outside that set (e.g., my staying at home while you go to Mexicali alone).

Using this model, we can see how it is possible to make a joint decision about what *you* will do or about what *I* will do: we can decide upon a shared option that substantively constraints

---

[4] Different theories of options will read the 'possible' in 'possible actions' differently: see Hedden (2012) for a review.
[5] To clarify: all this model says is that whenever we do something, that involves you doing something and me doing something. The model does not say that everything we can do is a sum of something I could do without you and something you could do without me: you and I can carry a heavy couch together, though neither of us would be able to do our part in this action without the other. Nor does it say that when we break down what we do into what you do and what I do, there will be any natural way of describing those individual components: there is likely no natural way to describe what I do when we dance a waltz other than to say that I am doing my part in our dancing a waltz. Thanks to Kieran Setiya for discussion on this point.

only one of our actions. Consider the set of possible combinations of our actions in which I gather the kindling for our campfire. Jointly deciding on this set would significantly constrain my actions, excluding from consideration any possible combinations of actions in which I don't gather any kindling. However, it would leave your actions largely unconstrained: any action you could perform that is compatible with my gathering the kindling is part of one of the combinations of actions in this set. Thus it seems that nothing of substance is missing if we describe our joint decision simply as a decision *that I will gather the kindling*, omitting the implicit clause 'and you will not prevent me from doing so.'[6]

The most obvious examples of joint deliberation concern joint action: where *we* will meet for lunch, whether *we* will have a child, whom *we* should hire. But on reflection, there is no barrier to our jointly deliberating about what *you* or *I* will do individually. This is a crucial point, as many of the speech acts I will argue are proposals in joint deliberation concern only one person's actions. Promises and offers concern only the speaker's actions; requests, demands, and commands concern only the addressee's. To deny the possibility of joint decisions like this, one would have to offer an alternative account of shared options that somehow excludes options that constrain only one agent's behavior while remaining flexible enough to capture the full range of possible joint activities. I don't see a principled way of doing this; nor do I see what would motivate the undertaking. The most natural account of shared options leaves room for joint decisions about what you or I will do alone.

1.2. Joint decisions

Our model of shared options entails an account of the content of joint decisions. Just as an individual's decision takes one of her options as its content, our joint decision will take one of our shared options as its content. So, the content of a joint decision will be a set of possible

---

[6] This clause seems to apply to promises: we would normally be affronted if someone accepted a promise and then prevented us from carrying it out. See Darwall (2011: 268-269) and Gilbert (2011b: 99).

combinations of our actions: e.g., the content of our joint decision that we will go to Mexicali is the set of possible action combinations in which we go to Mexicali.

But we want to understand not just the content of joint decisions, but their *force*. In the individual case, not just any representation of one's own actions counts as a decision: I can *suppose* that I will φ, *predict* that I will φ, or *tell a story* in which I φ. What distinguishes my *deciding* to φ from these other representations of my φing is its effects on my thought and behavior. If I decide to φ, then I should be disposed to take the means I believe to be necessary to φing, and avoid performing actions I believe to be incompatible with my φing (Bratman 1987: 15-17). I should also be disposed to take it for granted that I will φ in my subsequent deliberation, excluding options in which I do not φ from live consideration. These claims are not only descriptions of the psychological effects of decisions, but are also normative claims about practical rationality: if I decide to φ but fail to be disposed in these ways, then I am failing to conform to the norms of individual practical deliberation (cf. Bratman 2009).

Similarly, for a representation of a shared option to count as a joint *decision*, it must have implications for the deliberative participants' ongoing thought and behavior. If we jointly decide that I will φ, I should be disposed to take the means I believe to be necessary to φing and to avoid performing actions I believe to be incompatible with my φing. I should also be disposed to take it for granted that I will φ in my subsequent deliberation, excluding options in which I do not φ from live consideration.[7] These claims are not only descriptions of the psychological effects of joint decisions, but are also normative claims about the requirements of joint deliberation: if we jointly decide that I will φ, but I fail to be disposed in these ways, then I am failing to conform to the norms of our joint practical deliberation.

A joint decision, then, is a representation of a set of possible combinations of our actions that commits us to constraining our ongoing deliberation and action within its bounds.

---

[7] Later, I will appeal to the exclusionary effect that joint decisions have on subsequent deliberation to explain the similar exclusionary effects of promises, commands, and our other target speech acts. See Ch. 2, §2.2; Ch. 3, §1.1, §2, and §3.1; and Ch. 4, §2.

19

1.3. Shared reasons

Practical deliberation, whether individual or joint, is a *rational* process. Decisions are subject to normative standards: some decisions we ought to make, others we ought not. Akratic exceptions aside, when we deliberate about what to do, we aim to decide to do what we ought to do. In other words, we try to make a decision that is supported by our *reasons*. Similarly, when we deliberate together about what to do, we try to make a joint decision that is supported by our *shared reasons*. Whenever one extols the virtues of one's favored joint plan, or objects to a joint decision on the basis that it is unfair, inefficient or just a bad idea, one is appealing to shared reasons.

It is now customary to distinguish between *motivating* and *normative* reasons for action. A person's motivating reasons are the reasons that determine what decisions she *in fact* makes; her normative reasons are the reasons that determine what decisions she *ought* to make. Similarly for the joint case, we can distinguish between the reasons on the basis of which we *in fact* make our joint decisions, and the reasons that determine what joint decisions we *ought* to make. I will use 'shared reasons' to refer to the latter: our shared *normative* reasons. So, the question for our account of shared reasons is: what are the considerations we ought to take into account when we are deciding together what to do?

Here's a partial answer to this question: our shared reasons must be reasons *for each of us*. This claim is motivated by our second guiding idea – that participation in joint practical deliberation must be compatible with maintaining one's rationality and autonomy.

To fully participate in our joint practical deliberation, I must treat our joint decisions in the same way I treat my own individual decisions. When we jointly decide that I will φ, this should settle the question of whether I will φ directly: there is no further question to be asked, within my individual deliberation, of whether to φ. Thus when I φ, I should be acting *on our joint decision*, rather than acting on an individual decision that happens to align with our joint one.

This raises a worry. Autonomous agency involves acting on the basis of the beliefs, desires, values, and intentions that make up one's individual perspective of practical deliberation.

But I seem to be saying that joint practical deliberation requires us to give up our individual perspectives and base our actions on the perspective of the group instead. This sounds disturbingly like a call for individual subservience to the collective will.

The way to avoid these totalitarian overtones is to resist the idea that taking up the perspective of joint deliberation requires giving up one's individual deliberative perspective. We can render the two compatible by thinking of the former as part of the latter: the part of one's reasons that others share. Rather than thinking of joint practical deliberation as a separate process with its own *sui generis* set of reasons, we should think of our joint deliberation as occupying the intersection between our various individual perspectives of practical deliberation.

Less metaphorically, I am suggesting that we can make joint deliberation compatible with individual autonomy by requiring the shared reasons on which joint decisions are based to also be reasons for each participant individually. It is widely held that autonomy centrally involves acting on the basis of reasons you endorse.[8] So, if a joint decision is based on reasons that I can endorse within my individual perspective of practical deliberation, then I can act directly on that joint decision without sacrificing any autonomy. Conversely, if I *am* able to act on a joint decision autonomously, then it must be based on reasons that I *can* endorse as my own. Since the same reasoning applies to any participant in joint deliberation, our shared reasons must be reasons that *all of us* can endorse. Here's how I suggest we make this constraint precise:

UNANIMITY:

For any persons X, Y, Z, et al., proposition p, and action ϕ,

---

[8] For example, Frankfurt (1971) famously says that the reason why an unwilling addict is not autonomous is because she is acting for reasons she does not endorse (namely, her desire for the drug). Other theories of autonomy can be seen as pursuing the same idea, but offering different interpretations of what it means to endorse a reason. Frankfurt thought of endorsement in terms of higher-order desires: to endorse a reason for acting is to have a second-order desire that one act for that reason. In contrast, Watson (1975) argues that no higher-order desires are necessary: instead, the reasons one endorses are those that align with one's reflective judgments about what is good. Bratman (2007) proposes that the reasons one endorses are those one has standing, intention-like policies to treat as reasons within one's practical deliberation. We need not decide between these views here: they all agree that, in the central cases at least, autonomy requires acting only on reasons one believes to be genuine reasons for action. Thanks to an anonymous reviewer for asking the question that prompted this footnote.

21

(i) The fact that p is a shared reason for X, Y, Z, et al. in favor of X φing *only if*

the fact that p is a reason for each of X, Y, Z, et al. to prefer that X φ.

(ii) The *weight* of the shared reason for X, Y, Z, et al. provided by the fact that p in

favor of X's φing can be no greater than the weight of the reason provided by

the fact that p in favor of preferring that X φ for any of X, Y, Z, et al.

individually.[9]

UNANIMITY says that a consideration can be a shared reason for us only if it corresponds to

normative reasons of similar content and weight that apply to each of us individually.[10]

Two points of clarification are in order. First, UNANIMITY requires that our shared

reasons *for action* line up with our individual reasons *for preference*. This is simply because it doesn't

make sense to say that you have reason *for me to* φ. Your reasons have to bear on *your* actions or

attitudes, and my reasons have to bear on *my* actions or attitudes. To compare reasons across

persons, then, we need to look at reasons *for preference*, since I can have reason *to prefer* that you φ

and you can have reason *to prefer* that I ψ. The requirement thus is that p is a shared reason for us

in favor of my φing only if p is a reason for both you and me *to prefer* that I φ. 'Preference' here

must be read as *all-things-considered* preference: what one prefers in light of all of one's reasons. On

this interpretation, it would be akratic to say "I ought to φ, though I'd prefer not to." My reasons

for preference regarding my own actions are thus one and the same as my reasons for action; but

my preferences can also concern matters other than my own actions.[11]

---

[9] This second clause requires that we be able to compare the weight a reason has within our joint deliberative perspective with the reason it has in each of our separate individual perspectives. This raises tricky issues about whether, and how, it is possible to compare the weight of reasons across persons. Addressing these issues would take us too far afield, and so I will have to leave it as an undefended premise that interpersonal comparisons of weights are intelligible.

However, the intuitive idea should be clear enough. When we are deciding where to go to lunch, we should take the fact that you prefer Mexican food as a consideration in favor of going to Mexicali. However, this isn't as strong a consideration for us deciding together as it would be if you were deciding where to eat lunch alone. Deciding alone, each of us would give more weight to our own preferences; deciding together, we should give our individual preferences equal weight, as that is the common denominator between our two perspectives. This is the idea UNANIMITY is meant to capture. Thanks to Kieran Setiya for asking me about this issue.

[10] I defend UNANIMITY from some apparent counterexamples in §5.2 below. UNANIMITY plays an important role in generating the deliberative theory's predictions about when we have standing to make demands; see Ch. 4, §3.2.

[11] Here there is a contrast between reasons for action and reasons for attitudes such as belief and intention. I can have reason *to prefer that* I believe that God exists (since believing so would make me happier) without thereby having

Second, the 'individual reasons' with which UNANIMITY says our shared reasons must coincide need not be egoistic or self-regarding reasons. My individual reasons are simply the reasons that apply within my individual practical deliberation, including other-regarding and moral reasons as well as self-regarding ones. Thus UNANIMITY does *not* say that we have shared reason to make a joint decision only if that decision is in each of our *interests*.[12]

UNANIMITY provides a necessary condition for a reason to be shared; but I do not claim that it is sufficient. There might be other constraints a reason must meet to count as shared: perhaps our shared reasons must be common knowledge between us, or perhaps they must be related to our personal relationship or joint projects in the right way. But we need not decide these questions here. My aim in this chapter is not to answer every question about joint practical deliberation; it is to give an account that is substantive enough to predict and explain the features of our target speech acts. And for those purposes, UNANIMITY is all we need to know about shared reasons.

2. The norm of joint decisions

When an individual deliberates about what to do, she tries to make a decision that is justified by her reasons. In doing so, she guides her decisions by a *norm* that says what decisions are permissible given her reasons and options. Similarly, when we deliberate together about what to do, we must try to make a decision that is justified by our shared reasons. In doing so, we are holding our joint decisions up to a norm that says what joint decisions are permissible given our shared reasons and options. Let us dub this *the norm of joint decisions*. Call joint decisions that are permitted by this norm *warranted*, and those that are not *unwarranted*.

---

reason *to believe* that God exists. I can have reason *to prefer that* I form an intention to drink a toxin (because an eccentric billionaire will pay me money if I have this intention) without having reason *to intend* to drink the toxin (since the payment is not contingent on my actually drinking the toxin; see Kavka 1983). However, I do not think that we can similarly pull apart my reasons for preferring that I do an action from my reasons to do that action. If an eccentric billionaire offers to pay me money if I φ, this is not just a reason *to prefer that* I φ: it is also a reason *to* φ. Thanks to Susanna Rinard for discussion here.

[12] Thanks to Sarah Paul and Tamar Schapiro for pressing me to clarify this point.

What is the norm of joint decisions? Part of the answer is simple. An individual person's decision to φ is warranted just in case she has *sufficient reason* to φ: that is, there is no incompatible option ψ such that she has determinately more reason to ψ instead. For example, I am warranted in deciding to go to Mexicali just in case there's no other restaurant I have more reason to choose. Extending this idea to the joint case, we get:

> **SUFFICIENT SHARED REASON:** a joint decision that we will φ is warranted *only if*
> we have *sufficient shared reason* to φ, that is, there is no incompatible shared option ψ
> such that we have determinately more shared reason to ψ instead.

So: we are warranted in jointly deciding to go to Mexicali only if our shared reasons support this option at least as much as they support any other shared option (say, going to Clover).

However, a joint decision might meet the requirement of SUFFICIENT SHARED REASON and still fail to be warranted. Why? Because it could be that our shared reasons sufficiently support a joint decision that I will φ, and yet the reasons that apply within my individual, all-things-considered practical deliberation give me decisive reason *not* to φ. Though UNANIMITY implies that every shared reason is a reason for each of us, the converse entailment does not hold: not every reason that applies to each of us individually is a shared reason for us. (In fact, unless our individual reasons are identical, UNANIMITY implies that some of them must not be shared). There is thus no guarantee that our shared reasons will always favor the same actions that our individual reasons do. If you and I are writing a paper together, then perhaps our shared reasons will most support my spending the weekend working on our paper; when in fact, all things considered, I ought to spend the weekend with my family instead. If we could warrantedly decide that I will spend the weekend working on our paper, then I would have to choose between the action warranted within our joint practical deliberation (to work on our paper) and the action required by my individual reasons for action (to spend time with my family).

If joint practical deliberation is to be compatible with individual rationality (which was our second guiding idea), these kinds of conflicts cannot be possible. The norm of joint decisions

must include a proviso that prevents warranted joint decisions from conflicting with individual reasons in this way. I suggest the following:

> **JOINT-INDIVIDUAL COHERENCE:** a joint decision that we will φ is warranted *only if*, conditional on the joint decision's having been made, each of us has sufficient reason, all things considered, to do his or her part in our φing.

JOINT-INDIVIDUAL COHERENCE implies that a warranted joint decision will not require any person to do something she ought not do at the time the decision is made.[13] By building this requirement into the norm by which we evaluate joint decisions, we ensure that the results of our joint deliberation, if warranted, will respect each of our individual reasons for action.

Both SUFFICIENT SHARED REASON and JOINT-INDIVIDUAL COHERENCE give necessary conditions for a joint decision to be warranted. Whether meeting both of these conditions is sufficient for a joint decision to be warranted, or whether there are other criteria it must meet, will have to remain a question for further work.

---

[13] Two clarifications about the timing of JOINT-INDIVIDUAL COHERENCE. First, it may happen that a joint decision that satisfied JOINT-INDIVIDUAL COHERENCE when it was made still ought not be executed at a later time, if the situation has changed in relevant respects. In an uncertain world, such diachronic conflicts are inevitable, and will occur for individual decisions as well. The only way to avoid diachronic conflict would be to never plan for the future. We are interested instead in ruling out *synchronic* conflict between warranted joint and individual decisions.

Second, JOINT-INDIVIDUAL COHERENCE only requires that agents have sufficient reason to act in accordance with a joint decision *conditional on its being made*. Thus it does not require that agents have sufficient reason to act in accordance with a joint decision *before* it has been made, but instead only immediately *after* it is made. This distinction becomes important when we consider Prisoner's Dilemmas: scenarios in which each of us has decisive reason individually to 'defect' rather than 'cooperate,' though we would both be better off if we both cooperated than we would be if we both defected. A central function of joint decisions is to allow us to escape these scenarios by jointly deciding on the shared option in which we both cooperate. But it is the nature of Prisoner's Dilemmas that neither of us has sufficient reason to cooperate *prior to* our jointly deciding to do so. The joint decision to cooperate still satisfies JOINT-INDIVIDUAL COHERENCE, however, because once the joint decision is made, *then* we will each have sufficient reason to cooperate, since the fact that we have jointly decided to cooperate gives both of us additional reason to do so (and thus assurance that the other person will cooperate as well). Thanks to Jed Lewinsohn for discussion on this second point.

25

3. The process of joint practical deliberation

3.1. Two deliberative methods: propose-and-ratify and propose-and-challenge

We have now characterized the formal structure of joint practical deliberation; but we have yet to say how two or more persons can actually perform an activity that satisfies this formal description. How do we actually go about deciding together what to do?

To count as an activity of joint practical deliberation, a process must meet two criteria. First, for us to count as engaged in joint practical *deliberation* (rather than an arational decision-making process), the process by which we make joint decisions must be guided and constrained by the norm of joint decisions. Second, for us to count as engaged in *joint* practical deliberation (rather than each deliberating separately), the process by which we make joint decisions must be a *joint activity*: meaning, at minimum, that it must involve all of us playing some part.[14]

One procedure that meets these two criteria is what I call the *propose-and-ratify* method. Propose-and-ratify proceeds in two steps. First, a speaker proposes a potential joint decision that she takes to be warranted. Second, the addressees evaluate the proposed joint decision, judge whether it is warranted, and then decide whether to accept the proposal on that basis. A proposed joint decision comes into force just in case all participants explicitly accept it.

Propose-and-ratify is what first comes to mind when one thinks of joint decision-making. I say, "Shall we go to Mexicali?" You reply, "Sure!" And a joint decision is made. But it is important to see that propose-and-ratify is not the only way we can go about making joint decisions. By adopting the propose-and-ratify method, we implicitly make non-trivial choices about how to proceed in two important scenarios.

First, suppose a joint decision is proposed and then no further words are spoken. What is the result? On propose-and-ratify, the proposal is rejected by default: if the addressees do not explicitly accept the proposal, no joint decision is made. Call this the *initial default*: the result we get if a joint decision is proposed and then nothing else is said.

---

[14] As explained at the outset of this chapter, I want to remain neutral on what else is required for joint activity.

Second, suppose that you propose a joint decision that I disagree with. Arguing it out, we are unable to reach a consensus on whether to accept the proposed joint decision. What is the result? On propose-and-ratify, the proposal is again rejected by default: if all participants do not explicitly accept the proposal, no joint decision is made. Call this the *final default*: the result we get if a joint decision is proposed and then the participants cannot agree on whether to accept it.

Both the initial and final defaults of the propose-and-ratify method are to reject the proposed joint decision. In this way, the propose-and-ratify method is conservative: it errs on the side of preserving the status quo. But nothing in the nature of joint practical deliberation forces us to be conservative in this way: we could also adopt a more liberal method of deliberation in which the default result is for a proposal to be *accepted*. Call this second deliberative procedure the *propose-and-challenge* method.

Propose-and-challenge also begins with a speaker proposing a joint decision she takes to be warranted. But then, instead of explicitly ratifying the proposal if they think it is warranted, the addressees are expected to *challenge* a proposed joint decision if they think it is *un*warranted. When someone challenges a proposal, the other participants must then decide whether to accept the challenge: does the challenge show that the original proposal was unwarranted? If they think so, they accept the challenge; a challenge is successful just in case everyone explicitly accepts it. A proposed joint decision comes into force just in case there have been no successful challenges to it – meaning either that no challenges have been raised, or that any challenges that have been raised have failed to gain acceptance.

The initial default in propose-and-challenge is for a proposal to be accepted: if nothing is said after a proposal is made, the proposed joint decision comes into force by default. The *final* default in propose-and-challenge is also for a proposal to be accepted. A challenge blocks a proposed joint decision only if everyone accepts the challenge as warranted. So, while a propose-and-ratify proposal is accepted only if all participants explicitly accept it, a propose-and-challenge proposal is rejected only if all participants explicitly *reject* it. In other words, while propose-and-ratify resolves intractable disagreements by rejecting the proposed joint decision, propose-and-

27

challenge resolves intractable disagreements in favor of the proposed joint decision, bringing it into force. If we think of a deliberative proposal as a defendant in a criminal court, then propose-and-ratify presumes a proposal guilty until it is proven innocent, while propose-and-challenge presumes it innocent until proven guilty.[15]

Propose-and-challenge meets our two criteria for a process of joint practical deliberation. First, proposals accepted by the propose-and-challenge method are guided by the norm of joint decisions, since the participants are expected to challenge proposals they take to violate this norm. Challenges are also evaluated by the norm of joint decisions: the participants are expected to accept a challenge just in case they believe that it shows the proposal to be unwarranted. Propose-and-challenge also meets our second criterion. Each deliberative participant plays a role in the propose-and-challenge process by being prepared to raise any objections that come to mind. Even when a proposed joint decision is accepted with no further words, it is made jointly by all, since the addressees' silent acceptance signals that the proposal has passed a collective process of evaluation.[16]

Thus we have two very different procedures in our deliberative toolkit. Which method we ought to employ will depend upon our situation.

Two differences between our deliberative methods stand out as most significant. The first regards the two methods' different *propensities to error*. Since the propose-and-ratify method defaults to rejecting proposals, it makes it more difficult for proposals to come into force. As a result, the propose-and-ratify method protects against *false positive errors*: the error of accepting a proposed joint decision that is unwarranted. In contrast, the propose-and-challenge method

---

[15] In principle, it should be possible to have a method with different initial and final defaults: say, a proposal that initially defaults to acceptance (like propose-and-challenge), but in the case of disagreement defaults to rejection (like propose-and-ratify). However, I cannot think of any actual cases that fit this description. The reason, I suspect, is that the considerations that favor setting the initial default one way favor setting the final default the same way, and vice versa. When we have reason to be slow and careful, we should set both defaults to rejection (propose-and-ratify); when we need to be quick and decisive, we should set both defaults to acceptance (propose-and-challenge). The reason why we don't have 'hybrid' proposals may be just because we have no need for them.

[16] In the right context, however, even silence could count as a challenge (say, when combined with a pointed stare). What matters is not the difference between being silent and uttering words *per se*, but the difference between performing a communicative act that is reciprocally recognized as raising a challenge, and performing no such act. Thanks to Bradford Skow for pointing this out.

makes it easier for proposed joint decisions to come into force, as it defaults to acceptance. This makes propose-and-challenge more prone to false positive errors than propose-and-ratify. But it also means that propose-and-challenge is better than its counterpart at avoiding *false negative errors*: the error of rejecting a proposed joint decision that is in fact warranted. In short, our two methods make different trade-offs between accuracy and decisiveness: propose-and-ratify favors accuracy over decisiveness; propose-and-challenge favors decisiveness over accuracy.

This makes them appropriate in different situations. Propose-and-ratify is better when we want to deliberate carefully. If the stakes are high, and we have ample time to discuss the options, we should use propose-and-ratify, making sure a joint decision is made only if everyone explicitly accepts it. So, for example, two partners deciding whether to have a child should clearly make this decision using propose-and-ratify. In contrast, propose-and-challenge is better when we need to be quick and decisive. If the stakes are low, or the costs of taking more time to deliberate are high, then we should opt for propose-and-challenge. Suppose we are firefighters rescuing people from a burning building. Here it matters more that we decide on a plan, and decide on it quickly, than that we choose the best of all possible plans. So it makes sense to use propose-and-challenge: "You search the first floor, I'll go upstairs!"

The second central difference between our methods concerns the balance of power they create between the speaker and addressee(s). Propose-and-challenge gives the deliberative power to the speaker; propose-and-ratify gives it to the addressee(s). In propose-and-challenge, the speaker's proposals come into force automatically unless someone raises a convincing challenge to them. If a challenge convinces the speaker, then she can accept it and thereby reject the proposal. But since the final default is acceptance, then if the speaker is unconvinced by the challenges that are raised, her proposal will still go through. Thus propose-and-challenge gives the speaker the last word as to whether a proposal comes into force.

Propose-and-ratify, on the other hand, puts the power in the hand of the addressee(s). A proposal made by propose-and-ratify comes into force *if and only if* the addressee(s) explicitly accept it. Thus the addressee of a propose-and-ratify proposal has the power to veto it by

withholding her assent, or bring it into force by expressing her acceptance. Thus propose-and-ratify gives the addressee(s) the last word as to whether a proposal comes into force.

These different power dynamics play a crucial role in our decisions about which method to employ. As we shall see (Ch. 3, §1.3 and §3.3), the speech acts that use propose-and-ratify, offers and requests, are most appropriate when we want to give more deliberative power to the addressee. The speech acts that use propose-and-challenge — promises, commands, and demands — are most appropriate when we have reason to give more power to the speaker.

3.2. Good faith

Say we go through the deliberative motions: I say, "Shall we go to Mexicali?" and you say, "Sure!" Is that all it takes for us to count as having made a joint decision? Not exactly. If our acts of proposal and acceptance are to constitute the making of a joint decision, both of these acts must be performed *in good faith*.

To see what I mean, note that it is possible for two or more people to *appear* to make a joint decision without actually making one. Consider two actors who pretend to make a joint decision in a play. If one were to hold the other to their 'joint decision' afterwards, she would clearly be confused: no joint decision was made. The actors were not deciding together, but only *pretending* to do so.

Sometimes, the pretense is less public. One person may think she is engaged in joint deliberation while the other is only pretending to deliberate with her. To count as deliberating with you, I must evaluate potential joint decisions on the basis of our shared reasons, proposing and accepting only decisions that I take to be warranted by those reasons. Suppose instead that I pay no attention to our shared reasons and instead simply propose and accept joint decisions based on whether their acceptance will suit my purposes, whether or not they are justifiable in terms you could accept. If I behave in this way, I am not deliberating with you; I am only *pretending* to do so.

Let us call this kind of pretending *deliberating in bad faith*, and its opposite — genuine participation in joint deliberation — deliberating in *good faith*. When I deliberate in bad faith, I am not participating in joint practical deliberation at all: I am just pretending to do so. We are not deciding *together*, and so we cannot *decide* anything together. My failure to participate robs our interaction of the mutuality that is essential to joint decision-making. The idea of a joint decision implies a meeting of the minds; if I am deliberating in bad faith, no meeting of the minds occurs. My bad faith thus renders our joint deliberation impotent: so long as I am deliberating in bad faith, we cannot make any joint decisions — though we may falsely appear to do so.

How do we tell whether someone is deliberating in bad faith? The following seems to be a sufficient condition: if a person knows that the norms of joint practical deliberation require her to ϕ, and yet she intentionally does not ϕ, then she is deliberating in bad faith. Here ϕing is any move within the deliberative process: proposing, challenging, or ratifying a joint decision. If I propose a joint decision I know to be unwarranted, accept a joint decision I know to be unwarranted, or refrain from challenging a decision I know to be unwarranted, I am deliberating in bad faith.[17] I am not doing my part in our activity of making decisions together on the basis of reasons we share.

The requirements of good faith help assuage one *prima facie* worry about the propose-and-challenge method. You might think that propose-and-challenge allows me to impose any joint decision I like on you: all I have to do is propose the decision using propose-and-challenge and then refuse to accept any of your challenges. The good faith requirement means that I cannot do this. To deliberate in good faith, I must propose only those joint decisions that I believe to be warranted by our shared reasons. And if I believe a joint decision to be warranted, I must think it is supported by reasons that you can endorse (by UNANIMITY, §1.2) and that, if the joint decision is made, you will have sufficient reason to act in accordance with it (by JOINT-INDIVIDUAL

---

[17] Note that only proposing or accepting a joint decision one *knows to be* unwarranted counts as bad faith, not proposing or accepting an unwarranted joint decision *simpliciter*. I can propose or accept an unwarranted joint decision in good faith so long as I mistakenly believe that it is warranted.

COHERENCE, §2). I cannot propose in good faith any joint decision I take to violate these conditions. Moreover, if you raise a challenge to my proposed joint decision, then I must consider whether your challenge shows my proposal to be unwarranted, and base my response on that judgment. If I judge that your challenge shows my proposal to be unwarranted, then good faith requires me to accept your challenge. The only way I can legitimately reject your challenge and stand by my proposal is if I believe that it is warranted even after considering your objection. Thus while propose-and-challenge gives the speaker the last word, her 'last word' must be a judgment made in good faith about whether her proposal is justifiable to her addressee(s). You cannot use the propose-and-challenge method to unilaterally impose your will on others because unilateral imposition is a form of deliberative bad faith.

The requirements of good faith will also play a central role in explaining why our target speech acts are invalid under conditions of coercion or deception. This effect is most salient with promises. If I get you to promise by making a coercive threat, or by deceiving you on some important matter of fact, then your promise does not bind you in the normal way. In Chapter 2 (§4), I argue that this is because coercion and deception are both forms of deliberative bad faith.[18]

3.3. Retraction

People can change their minds. Having decided to go to the library, but then seeing that it is a beautiful day, I might abandon my earlier plan and decide to work in an outdoor café instead. I thereby retract my earlier decision. This reverses the decision's effects on my thought and behavior. Once I've retracted my decision to go to the library, I will no longer be disposed to go there, or rationally criticizable if I do not, and I can now freely consider options that are incompatible with going to the library (such as going to an outdoor café).

---

[18] Arthur Ripstein offers a similar explanation of why coerced and deceived consent are invalid: "In cases of force and fraud, consent fails for lack of a united will. Both fraud and force stop the parties to an agreement from uniting their wills because the person committing either is already unilaterally determining how the other's means will be used" (2009: 131).

Similarly, we can change our minds together, retracting a joint decision we previously made. As in the individual case, retracting a joint decision reverses its psychological and normative effects. Once we retract our decision to go to Mexicali for lunch, neither of us is criticizable if we do not go there, and we can now consider shared options that are incompatible with our earlier decision (such as going to Clover instead).

How do we retract a standing joint decision? In the same way we make new decisions. One of us can propose to retract our standing joint decision; then, the others will assess whether the retraction proposal is warranted and signal their acceptance or rejection on that basis. If the proposal is accepted, then the joint decision is thereby retracted, expanding the shared options to include some of the possible action combinations that the decision had previously ruled out.[19]

Just as we can propose new decisions by means of either propose-and-challenge or propose-and-ratify, we can also propose to retract decisions by either of these two methods. And so a person who wants to propose retraction faces a choice: use propose-and-challenge, thus making the default for the decision to be retracted, or use propose-and-ratify, making the default for the decision to stay in force? The same considerations reviewed in §3.1 apply to this choice. When one has reason to err on the side of the status quo, or to give the addressee more deliberative power, one should propose retraction using propose-and-ratify. But when there is less reason to be cautious about retracting a decision, or reason to take deliberative power into one's own hands, then one can propose retraction by means of propose-and-challenge.[20]

4. Interim summary

The account of joint practical deliberation I have proposed is as follows. Joint practical deliberation is a process of making a joint decision about which of our shared options to perform on the basis of our shared reasons. Our shared options are mutually exclusive sets of possible

---

[19] The question of *which* actions are added back in to the shared options is an instance of the *problem of permission* (Lewis 1979a; Yablo 2009). I won't try to address this notoriously difficult question here.

[20] These considerations play a large role in my explanation of why the addressee of a promise can release the promisor from her obligation, while the promisor cannot release herself. See Ch. 2, §1.5 and §3.

combinations of our actions (§1.1). We can make joint decisions concerning only one of our actions by deciding on a shared option that constrains only that person's actions. When we make a joint decision, we select one of our shared options and commit ourselves to constraining our subsequent deliberation and action within its bounds (§1.2). Our shared reasons are the considerations that we ought to take into account when deciding between our shared options. I argued that such considerations must be reasons for each of us: a fact can only be a shared reason *for us* to φ if it is a reason *for me* to prefer that we φ and a reason *for you* to prefer that we φ (UNANIMITY, §1.2).

Any activity of joint decision-making must be guided and constrained by a normative standard that I called the norm of joint decisions (§2). When we deliberate together, we try to ensure that the joint decisions we make are warranted by the lights of this norm. I proposed two necessary conditions for a joint decision to be warranted: it must be supported by our shared reasons at least as much as any other decision available to us (SUFFICIENT SHARED REASON) and, conditional on our making the joint decision, we must each have sufficient reason, all-things-considered, to act in accordance with it (JOINT-INDIVIDUAL COHERENCE).

We can evaluate potential joint decisions using two different deliberative procedures (§3.1). In the propose-and-ratify method, a proposed joint decision is rejected by default unless all participants explicitly ratify it. In the propose-and-challenge method, a proposed joint decision is accepted by default unless all participants accept some challenge to it. These methods have different costs and benefits that make them appropriate in different situations. Propose-and-ratify is more careful and puts more power in the hands of the addressee(s); propose-and-challenge is more decisive and gives more power to the speaker. For either of these methods to result in a binding joint decision, however, all parties must participate in good faith (§3.2): they must base their proposals, acceptances, and challenges on their actual judgments about whether the joint decision in question is warranted by their shared reasons. Finally, just as we can make new joint decisions by these deliberative methods, we can also retract our prior joint decisions, thereby opening up shared options that were previously closed (§3.3).

5. Deciding together in everyday life

We arrived at our account of joint practical deliberation by unpacking the implications of two abstract ideas: the idea that joint deliberation is analogous in structure to individual deliberation, and the idea that joint deliberation must be compatible with individual rationality and autonomy. The account can be read as a description of what it would take for two or more persons to rationally and autonomously engage in a single activity of the very same kind as the activity an individual person undertakes when she decides what to do.

However, one might doubt that this abstract description actually captures our concrete practices of everyday joint decision-making. For all I have said so far, it could turn out that we never actually perform the activity I have called 'joint practical deliberation'. Perhaps we make decisions together in a way that is less analogous to individual practical deliberation, or less scrupulous about respecting rationality and autonomy, than I have suggested. In short, one might worry that the theory just presented fails to capture the way we actually make joint decisions in everyday life.

I will consider two objections along these lines. The first charges that my account cannot accommodate the possibility of *compromise*. The second objection targets the UNANIMITY constraint on shared reasons I defended in §1.2.

5.1. The possibility of compromise

We often manage to make joint decisions in the face of deep disagreements about what is or ought to be the case. When things go well, such disagreements end in compromise: the participants agree upon a joint decision that neither would have chosen if she had sole control. Each person would prefer that a different decision were made, but accepts the agreement because she can't get the other participants to agree to the decision that she thinks is best.

This everyday reality seems to be in tension with the good faith requirement I defended in §3.2. There I claimed that, to deliberate in good faith, one must accept only joint decisions one believes to be warranted. To accept a joint decision one believes to be unwarranted is to ignore

35

the shared norms and reasons that one must attend to in order to count as deliberating at all. "But," our objector might say, "People accept joint decisions they believe to be unwarranted *all the time!* That's what compromise *is*: accepting one agreement when you think another would be better. Your account implies that compromises are acts of bad faith, and therefore that the resulting joint decisions are not valid. But this could not be further from the truth. Far from being an act of bad faith, we take the willingness to compromise to be an essential deliberative virtue. And compromises are no less binding than any other agreements."[21]

Where this line of objection goes wrong is in its characterization of compromise: in particular, the claim that compromise involves accepting a joint decision one believes to be unwarranted. Once we take into account all of the reasons relevant to a compromise – including, crucially, the fact of disagreement itself – we see that compromises can be, and often are, the optimal choice by everyone's lights.

Consider two examples of everyday compromise:

MOVIES: My partner Dee Dee and I are choosing a movie to watch together on a weekend night. Dee Dee likes funny, happy movies; I like mindless action movies. If Dee Dee had her choice, we would watch a musical or a slapstick comedy; if I had my choice, we would watch a spy thriller or a blockbuster superhero movie. But we want to watch a movie together, so we compromise and watch *Beverly Hills Cop*, a buddy cop movie that is lighthearted and funny (for Dee Dee) but also has some action (for me).

SHOPPING: My friend Stephen and I are planning a 4th of July barbeque, and are shopping for food in preparation. I think that people tend to eat a lot of food at these things, while Stephen thinks that we always buy too much food. I grab four packages of sandwich rolls off the shelves. Stephen says, "That's way too much! We only need two." I say, "What if we run out of food?" Stephen sighs. "Fine – we can buy three, but no more." I put one package of rolls back, and we move on.

---

[21] Thanks to Sally Haslanger for raising this objection and putting it in a particularly compelling form.

Here's how our objector would describe these cases. In MOVIES, Dee Dee and I disagree about what movie we ought to watch. I think the best choice would be *Die Hard*; she thinks we should watch *Singing in the Rain. Beverly Hills Cop* is neither her top choice nor mine: so when we agree to watch it, we are each accepting a joint decision that we think is unwarranted.

In SHOPPING, the objector would say, Stephen thinks we ought to buy two packages of sandwich rolls, while I think we ought to buy four. Neither of us thinks that three packages is the right number. Thus when we agree to buy three packages, Stephen and I each accept a joint decision we take to be unwarranted.

To the contrary, I think that in both examples, both of the participants believe the compromise decision to be warranted. The objector's descriptions miss some essential details.

Start with MOVIES. Dee Dee's and my conflicting preferences concern what kind of movie we would each like to watch *if* we were watching a movie alone. But we want to watch a movie *together*, and neither of us wants the other to be miserable throughout the experience. Thus while I would want to watch *Die Hard* if I were alone, I don't want to watch *Die Hard* with Dee Dee unhappy and bored by my side; the same is true for Dee Dee vis-à-vis *Singing in the Rain*. When we take into account the fact that we want to watch a movie together that both of us will enjoy, then *Beverly Hills Cop* stands out as the clear best choice. We both realize this, and so can jointly decide to watch *Beverly Hills Cop* in good faith.

In SHOPPING, Stephen and I don't have conflicting preferences: we each want to buy enough food for the party without spending more money than is necessary. Instead, we have conflicting beliefs. Stephen believes that our guests will only eat two packages of sandwich rolls, while I believe that they are likely to eat four packages. Thus it seems plausible to say, with the objector, that neither of us believes three to be the best choice.

However, Stephen and I don't only want to make the optimal trade-off between price and quantity of food; we also want to make our shopping decisions quickly and efficiently. If we took the time to try to convince each other of our views about how much we need of every item, our grocery shopping experience would be drawn-out and painful. Even if we would save a couple

dollars, it wouldn't be worth the headache. Once we add this third value to the equation, the balance of reasons changes: we are both willing to accept some error in the quantity/price trade-off in exchange for a faster and smoother deliberative process. This justifies our compromise: by splitting the difference between our best guesses, we manage to quickly arrive at a decision that neither of us thinks is egregiously incorrect. Stephen and I would both rather decide *now* to buy three packages of sandwich rolls than argue further about how much our guests will eat. And so we can both accept this compromise in good faith.

Our examples illustrate two general reasons for compromise. First, we often want to do something *together* on terms that are mutually agreeable, as when Dee Dee and I want to watch a movie that neither of us will hate. To do something together rather than separately, we often have to compromise, agreeing on an activity that we are both willing to do even if neither of us would have chosen it alone. When we do this, our joint decision is still warranted and made in good faith, because we care more about finding an activity we can do together than we care about what particular activity we do.

Second, we often compromise simply because it is not worth the effort to argue further about what would be best. An analogue of this happens in individual deliberation: we sometimes make a decision in full knowledge that if we deliberated longer and more carefully, we would likely make a different and better choice. This is because the value that we might gain by deliberating more carefully is outweighed by the costs of spending more time deliberating. A quick-and-dirty decision is often justified by the fact that it is *quick*. In joint deliberation, these considerations of cognitive efficiency favor compromise. The disagreement that leads to a compromise might be such that we could in principle reach an agreement if we took the time to talk it through. (Stephen and I could sit down with our guest list, guess how much each individual is likely to eat, and calculate a much more precise estimate). But often we would each rather make a compromise decision *now* than argue it out until we reach consensus. It is this shared desire for cognitive efficiency that justifies our compromise.

I conclude that our account of joint deliberation does not preclude the possibility of compromise. Compromise does not require one to accept joint decisions one believes to be unwarranted. Compromise instead requires a sensitivity to certain considerations that warrant joint decisions that one might not have preferred on other grounds – such as the value of deliberating efficiently and acting together.

5.2. A challenge to UNANIMITY

The second objection we will consider targets UNANIMITY, the requirement that the reasons on the basis of which we make joint decisions must also be reasons for each of us individually (§1.2). This requirement seems at odds with some paradigm cases of joint decision-making. When a legislature passes a law by a narrow majority, the reasons on which the law is based (e.g., the fact that it will lower carbon emissions) are bound not to be accepted by a large number of the legislature's members. A hiring committee might choose to hire a candidate for the reason that she specializes in Kant, even if some of the committee members think that the department does not need a Kant scholar. An admissions committee might admit a student for the reason that she is a legacy (descended from alumni), even though some of its members think that legacy considerations should not be taken into account.

UNANIMITY implies that the people in these examples must be making some sort of mistake. They must either be incorrect about what they have reason to prefer individually, or be unjustifiably making a joint decision on the basis of a reason that is not shared. But intuitively, these examples do not seem to involve any such mistake. To the contrary, they seem to be central cases of well-functioning joint deliberation. Why can't it be the case both that *I* personally have no reason to prefer that we hire a Kant scholar and that, in spite of this, *we* justifiably decide to hire someone on the basis that she is a Kant scholar?[22]

---

[22] Thanks to Bradford Skow for helping me clarify the way in which these examples are in tension with UNANIMITY.

On the basis of examples like these, Michael Bratman argues that, contra UNANIMITY, the considerations on which joint decisions are based need not be accepted by all parties to those decisions (2014: 132-150). Instead, he suggests, all that is required for us to share a reason is that we have a shared commitment to treating it *as* a reason in our joint deliberation:

> Sharing a commitment to certain weights seems closer to a kind of shared intention than to a common value judgment ... In participating in such a shared intention concerning weights, each person will normally have some sort of supporting evaluative judgment. But such background judgments need not interpersonally converge, and need not strictly correspond to what is favored by the shared intentions concerning weights (137).

So, though not all members of the admissions committee think that the fact that a student is a legacy is a reason in favor of admitting her, they have accepted a shared policy of treating this fact *as* such a reason in their deliberations. And, Bratman submits, this is all it takes for legacy considerations to count as a genuine shared reason within their joint deliberation.

Bratman's diagnosis points the way to my response. I agree that, in these cases, the participants have adopted a joint policy about how to arrive at their decisions. But these joint policies themselves must be justified by appeal to a set of shared reasons; and these underlying shared reasons are plausibly regarded as satisfying the UNANIMITY constraint.

To see the idea, consider a different example:

CHORES: Anna, Bert, and Cora are roommates, and are discussing how to divide up household chores. Nobody wants to clean the toilet. So, they decide to determine who cleans the toilet by drawing straws: they take three straws, cut one short, and draw them blindly. Whoever gets the short straw has to clean the toilet: this time, it's Anna.

Anna, Bert, and Cora decide who will clean the toilet on the basis of who draws the short straw. So, they seem to treat the fact that Anna drew the short straw as a shared reason in favor of her cleaning the toilet. But none of them (Anna least of all) takes the fact that Anna drew the short straw to be a reason *to prefer that* Anna clean the toilet. Is this a counterexample to UNANIMITY?

No. Note that the fact that Anna drew the short straw, taken out of context, does not count in favor of her cleaning the toilet. Facts about who drew what straw have no intrinsic relevance to the question of who should clean the toilet. It is only when we learn that Anna, Bert, and Cora *already decided* that whoever drew the short straw will clean the toilet that the fact that Anna drew the short straw becomes a relevant consideration at all. Anna, Bert, and Cora all have reason to prefer that they adopt some quick and fair procedure for determining who gets the undesirable chore. This unanimously shared reason justifies their decision to determine who will clean the toilet by drawing straws. In turn, the shared reason that justifies their deciding that Anna will clean the toilet is the fact that they jointly decided that whoever drew the short straw would clean the toilet, *plus* the fact that unlucky Anna drew the short straw. All of them, including Anna, take these facts together as a reason to prefer that Anna cleans the toilet. As little as Anna wants to do this chore, she *does* want to honor her agreement with her roommates. Thus the shared reasons in CHORES conform to UNANIMITY after all.

We can apply the same line of argument to Bratman's examples. Focus on the admissions committee case. Just as Anna et al. did not take the fact that Anna drew the short straw, on its own, to be a reason to prefer that she clean the toilet, the members of the admissions committee do not all take the fact that a student is a legacy, on its own, to be a reason to prefer that she is admitted. But, given the number of applications they have to get through, the committee members have a unanimously shared reason to prefer that they settle on a decision-making procedure that includes a clear set of admissions criteria. This deliberative procedure is itself likely to be a compromise. Though one member might have opted not to favor legacies if she were deciding alone, she might accept a shared policy of favoring legacies because she wants to settle on criteria that everyone can agree upon within a reasonable amount of time.

Once the committee settles on a procedure, they have unanimous reason to prefer that they decide in accordance with it: both for the sake of deliberative efficiency, and because they all care about honoring their joint commitments. Thus the shared reason that justifies the committee's decision to admit the legacy student is not the mere fact that she is a legacy: it is the

fact that she is a legacy *plus* the fact that the committee has jointly decided to use admissions criteria that give preference to legacy students. All the members of the committee, even those who are skeptical about favoring legacies, take these combined facts as a reason in favor of admitting the student. After all, even the legacy skeptics prefer following the committee's streamlined decision procedure to arguing *ad nauseam* about whether legacy students should be favored. Thus the admissions committee's shared reasons conform to UNANIMITY after all.

I believe the same reasoning applies, *mutatis mutandis*, to the other examples Bratman offers as challenges to UNANIMITY. It is telling that almost all of these examples involve decision-making in formalized, institutional contexts: hiring and admissions committees, legislatures, juries, and so on. (The few non-institutional cases Bratman cites can, I believe, be explained as straightforward compromises). Bratman is right that, in these contexts, we often make decisions on the basis of criteria that we do not take to be reasons in other contexts. But that is because we are engaged in a kind of elaborate drawing of straws: our joint decisions are guided by an agreed-upon formal procedure that sets out our decision-making criteria (e.g., legacy considerations) and methods for resolving disputes (e.g., counting votes, drawing straws, etc.). Immersed in a procedure of this kind, it is easy to think that the facts that the procedure tells us to treat as reasons are doing all of the justificatory work: e.g., that our shared reason for admitting the student is just that she is a legacy. But that misses a crucial part of the justification for our decision: our commitment to the procedure itself. If we keep the full set of our shared reasons in view, the challenge to UNANIMITY dissipates. Our shared reasons are indeed reasons for us all.

6. The deliberative theory

We have done the groundwork. Now we can state the theory that the remainder of this dissertation will be spent defending.

We begin from a salient difference between joint and individual practical deliberation: while individual deliberation can be performed in solitary thought, joint deliberation needs to happen in conversation. We need to *communicate* with one another to deliberate together. This is

why both of our deliberative methods start with a communicative act: the speech act of proposing a joint decision. This speech act will be the opening move in any process of joint decision-making. Given how often we need to decide together what to do, we probably perform this speech act quite often. So, the speech act of proposing a joint decision should be quite familiar — so familiar, in fact, that we should expect to find a word for it in ordinary language. And since there are multiple different ways of proposing joint decisions, we might expect our language to distinguish between them, giving different names to different types of proposals. This suggests that *several* recognizable everyday speech acts may be proposals in joint practical deliberation.

I submit that ordinary English distinguishes between such proposals on the basis of their *content* and their *intended method of evaluation*. First, we distinguish between proposals to make joint decisions concerning only the speaker's actions ("I will ϕ"), those concerning only the addressee's actions ("you will ϕ"), and those concerning multiple participants' actions ("we will ϕ"). Second, we distinguish between proposals that are meant to be evaluated by propose-and-challenge and those meant to be evaluated by propose-and-ratify. My hypothesis is that our everyday speech act categories map on to these distinctions as follows:

	Propose-and-challenge	Propose-and-ratify
"I will ϕ"	Promises	Offers
"You will ϕ"	Commands & Demands	Requests
"We will ϕ"	Agreements	

Promises and *offers* propose joint decisions that constrain only the speaker's actions: "*I* will ϕ." They are distinguished by their methods of evaluation: promises employ the propose-and-challenge method, while offers use the propose-and-ratify method. This captures the fact that

offers require explicit acceptance to bind the speaker to perform, while promises do not. If I offer to drive you home, and you do not explicitly accept my offer, then I am under no obligation to drive you home. But if I *promise* to drive you home, you don't need to say anything for my obligation to come into effect: so long as you do not object, I am bound (cf. Robins 1984: 101; Darwall 2011: 273).

Commands, demands, and *requests* propose joint decisions that constrain only the addressee's actions: "*You* will φ." Commands and demands do so by means of the propose-and-challenge method, while requests use propose-and-ratify. This reflects a difference in uptake conditions that parallels the difference between promises and offers. If I ask you to read my draft, but you do not explicitly accept my request, then you are not obligated to read my draft. But if a professor directs her students to read the first book of Plato's *Republic*, they need not explicitly accept her command for it to come into force.

Why do I put commands and demands in the same box? Because, as I argue in Chapter 4, these speech acts both propose joint decisions regarding the addressee's actions by means of the propose-and-challenge method. What distinguishes the two is the way in which their proposed joint decisions are justified. Demands propose what I call *commissive* joint decisions: joint decisions that commit the addressee to doing what she already had decisive reason to do. Commands propose what I call *enactive* joint decisions: joint decisions that decide for the addressee which of multiple permissible options she will perform. This distinction captures the intuitive difference between commands and demands: commands create new obligations, while demands hold people to their existing obligations.

Agreements are joint decisions that constrain the actions of both the speaker *and* the addressee(s): "We will φ." Agreements have two features that make them stand out on our table. First, unlike our other terms, 'agreement' is a success term. While it is perfectly possible to have a rejected promise, the phrase 'rejected agreement' is oxymoronic. If the parties didn't agree to it, it wasn't an agreement. So while our other terms refer to proposed joint decisions, the term 'agreement' refers to accepted joint decisions. Thus agreements are not speech acts, strictly

speaking; *proposals* to make agreements are speech acts. Second, we don't have words in English that distinguish between the propose-and-challenge and propose-and-ratify methods of proposing agreements. But the distinction still applies: consider the difference between "Let's go on a walk!" and "Would you like to go on a walk?" The former proposal is meant to be evaluated using propose-and-challenge, while the latter is meant to be evaluated using propose-and-ratify.

We can now state the key theses of this dissertation:

The Deliberative Theory of Promises: For S to promise A that she will ϕ *just is* for S to propose to A, by means of the propose-and-challenge method, that they make a joint decision to the effect that S will ϕ.

The Deliberative Theory of Offers: For S to make an offer to A that she will ϕ *just is* for S to propose to A, by means of the propose-and-ratify method, that they make a joint decision to the effect that S will ϕ.

The Deliberative Theory of Agreements: For S and A to make an agreement that they will ϕ *just is* for S and A to make a joint decision to the effect that they will ϕ.

The Deliberative Theory of Requests: For S to request of A that she ϕ *just is* for S to propose to A, by means of the propose-and-ratify method, that they make a joint decision to the effect that A will ϕ.

The Deliberative Theory of Demands: For S to demand that A ϕ *just is* for S to propose to A, by means of the propose-and-challenge method, that they make a commissive joint decision to the effect that A will ϕ.

The Deliberative Theory of Commands: For S to command A to ϕ *just is* for S to propose to A, by means of the propose-and-challenge method, that they make an enactive joint decision to the effect that A will ϕ.

It is time to begin defending these theses. We begin, in the next chapter, with promises.

Chapter 2

Promises

Promises have captured the imagination of moral philosophers for centuries. It is easy to see why. I can promise you that I will do almost anything – make you a cup of coffee, run a marathon, read *Anna Karenina*, abstain from chocolate for a month – and *voila!* I am suddenly obligated to do it. Promises seem to create new obligations out of thin air. Promises thus make vivid the question with which I began this dissertation: how is it possible to change what morality requires of us simply by declaring it to be so? This is the puzzle that famously led Hume to proclaim promising "one of the most mysterious and incomprehensible operations that can possibly be imagin'd" (1739/1978: 524).

In the nearly three hundred years that have passed since Hume's proclamation, a sizeable literature on promises has emerged. This makes promises an ideal testing ground for the deliberative theory. The detail in which moral philosophers have characterized promises' intuitive features gives us a wealth of data against which to test the deliberative theory's predictions. And, by comparing the deliberative theory to the alternative accounts of promises in the literature, we can bring its theoretical advantages into stark relief.

In this chapter, I will argue that promises are proposals in joint practical deliberation. More precisely, I will defend

> **The Deliberative Theory of Promises:** For S to promise A that she will φ *just*
>
> *is* for S to propose to A, by means of the propose-and-challenge method, that they
>
> make a joint decision to the effect that S will φ.[1]

---

[1] I am not the first to propose this view. Margaret Gilbert has defended a *"joint decision account"* of promises: for one person to make a promise to another is for them jointly to commit themselves...to the decision that one of them ('the promisor') is to perform one or more specified actions" (Gilbert 2011b: 99). However, the arguments Gilbert offers for this view are very different from the argument I will give in this paper. Gilbert's main argument for the joint decision view is that it vindicates a purported analytic connection between promises and obligation: that the proposition that I have promised to φ entails that I am obligated to φ. But this thesis is controversial: I am convinced neither that the thesis is true nor that the deliberative theory entails it (see §1.1). Thus I do not think we should be convinced by Gilbert's arguments for the joint decision view. To see the idea's true potential, we need to take another look.

My argument for this thesis will take the form sketched in the Introduction to this dissertation. The theory of joint practical deliberation offered in Chapter 1 gives us the resources to predict the properties that various proposals within this activity will have. If we take a careful look at these speech acts, we will make a surprising discovery: one of these speech acts has *exactly the same properties as promising*. A certain kind of proposal to make a joint decision regarding one's own actions turns out to have the very same normative effects, under the very same conditions, as a promise. Our target proposals *bind* in the same way as promises: by bringing a joint decision into force, these proposals bind the speaker to perform a certain action (§1.1), obligating her *to her addressee* in particular (§1.2) and constraining her future deliberation to accord with this decision (§1.3). As with promises, the addressee alone has the power to release the speaker from the obligation created by these proposals, by agreeing to retract the joint decision that they brought into force (§1.5). And like promises, these proposals only have their force when they secure uptake from their addressee (§1.4) and are not elicited by means of coercion or deception (§2). I submit that this cannot be a coincidence: the only viable explanation of these observations seems to be that promises are identical to their analogues in joint practical deliberation.

Sections 1 and 2 of the chapter spell out this argument in detail. Section 3 addresses an objection, and section 4 situates the deliberative theory in relation to the theories of promising in the extant literature.

The chapter concludes, in section 5, with a general challenge to alternative theories of promising. If the arguments of sections 1 and 2 are sound, then joint practical deliberation necessarily includes a speech act with the very same normative features as promising. Anyone who denies the deliberative theory is committed to holding that this speech act is not identical to promising. But this implies that promising is *redundant*: everything we do with promises, we could do instead with proposals in joint practical deliberation. This undercuts the motivation for alternative theories: why posit a further, *sui generis* speech act of promising if joint practical deliberation already contains a speech act with identical properties? Better to accept that promises and the relevant kind of proposals in joint practical deliberation are one and the same.

1. Similarities between promises and proposals in joint practical deliberation

My thesis is that promises are proposals to make joint decisions regarding one's own actions using the propose-and-challenge method. While we evaluate this thesis, however, it will be helpful to have a separate term for the relevant kind of proposal, so as not to beg the question regarding its relation to promises. Let us coin the term *I-proposal* to refer to a proposal to make a joint decision regarding one's own actions ("*I* will φ") by means of propose-and-challenge. This allows us to state the deliberative theory more succinctly: promises are I-proposals. My argument for this thesis is founded on the claim that I-proposals have the same properties as promises. In this and the next section I defend this claim. This section surveys five of promises' most important features and argues that I-proposals have each of these features as well.

1.1. Bindingness

Promises create obligations. When I promise you that I will read your paper, I thereby become *obligated* to read your paper. However, while promises typically generate obligations, this connection is defeasible (*pace* Gilbert 2011b: see footnote 3). It is sometimes permissible to break a promise: if I promise to come to your talk, but then come across an injured person who urgently needs to be taken to the hospital, I plausibly ought to break my promise and help the person instead. More carefully, then, we can say that when I promise to φ under normal conditions,[2] I come to have a strong *pro tanto* reason to φ that typically, though defeasibly, obligates me to φ.

Are I-proposals similarly binding? I think so. Suppose that you and I jointly decide to read each other's papers. This joint decision plausibly obligates each of us to read the other's paper. Suppose you carefully read and give detailed comments on my paper, and come to our meeting to find that I have not even glanced at your paper. You would rightly feel wronged. (See §1.2 below for more on this 'directed' nature of the obligation to keep promises and joint decisions). So, when we make a joint decision, we each become obligated to do our part in seeing

---

[2] By 'under normal conditions' I mean that (a) the promisee does not reject the promise and (b) no coercion or deception is involved. Please take this qualification as understood unless context indicates otherwise.

it through. As with promises, this obligation can be overridden − if enough is at stake, it can be permissible for me to violate our joint plan − but even when it is, the fact that we made a joint decision is a strong *pro tanto* reason in favor of my carrying it out.

If you grant that the above joint decision is binding, then it is hard to deny that joint decisions made by means of I-proposals are too. The fact that a joint decision concerns only the speaker's actions, or that it was proposed by means of the propose-and-challenge method, does not seem to diminish its moral relevance. So, if we jointly decide just that *I* will read your paper, and we decide this by means of propose-and-challenge, this joint decision will similarly obligate me to you to read your paper. Thus I-proposals, like promises, (defeasibly) generate obligations that are owed to their addressees. If we take on the deliberative theory, we can thus explain the normative force of promises by appeal to the normative force of joint decisions.[3]

The obvious next question is: why should we abide by our joint decisions? I will not endorse any particular answer to this question. Instead, I will show how the most popular accounts of promises' normative force can also be used to vindicate the normative force of joint decisions. Thus there is no principled basis for accepting the normative force of promises while denying the normative force of I-proposals. In short: *if promises are binding, then I-proposals are too.*

Consider first Rawls's conventionalist account: we have reason to keep our promises because to break a promise would be to unfairly free-ride on the beneficial promising practice, violating the principle of fairness (1971: 343; for more on conventionalism, see §4.1). The principle of fairness applies to joint decisions as well: to make a joint decision and then violate it is to free-ride, helping oneself to the benefits of joint deliberation without doing one's part in the practice by obeying its rules. Thus the principle of fairness equally plausibly gives us reason to abide by our joint decisions.

Second, consider the style of account offered by two-tier theories such as contractualism and rule consequentialism. On these accounts, the principle requiring us to keep our promises is

---

[3] You might be worrying that the obligation to abide by a joint decision seems *weaker* than the obligation to keep a promise. This objection is important enough to get its own section (§3); until then, let us set it aside.

a basic axiom of our moral theory, not reducible to any more general principle. But this principle itself is justified by appeal to our interests: our interest in the principle's being generally followed (Scanlon 1998; Hooker 2011) or our 'normative interest' in its being true (Owens 2012). (I discuss this style of account further in §4.3 of this chapter). Along similar lines, we might posit a principle requiring us to abide by our joint decisions as a basic axiom of our moral theory. We could then offer the very same two-tier justifications for this joint decision principle. We surely have an interest in this principle's being generally followed. And if we do have a normative interest in the promissory principle's being true, I cannot see why we would not have a similar interest in the truth of the joint decision principle.

Third, consider a Kantian account. Kant's Formula of Humanity enjoins us to "so act that you use humanity...always at the same time as an end, never merely as a means" (Kant 1785/1996: 80). On an influential interpretation, treating a person as an end in herself requires acting in a way she could coherently *agree* to (O'Neill 1985; Korsgaard 1986; Langton 1992). On this view, it is wrong for me to break my promise to you because you could not coherently agree to my doing so. For, by agreeing to my breaking a promise, you would thereby *release* me from the promise, and thus make it the case that I am not breaking a promise after all. Your agreement would thus be self-defeating: the action you are agreeing to – breaking my promise – is of a kind that can only be performed in the absence of your agreement. This argument applies equally well to joint decisions. Consider whether you could coherently agree to my violating our joint decision. By the same reasoning, this agreement would be self-defeating: by agreeing to my violating our joint decision, you would thereby *retract* that joint decision, and thus make it the case that my action would not violate any joint decision after all. *Violating our joint decision* is the kind of action that can only be performed in the absence of your agreement. Thus the Kantian can also say that my violating our joint decision is wrong because doing so treats you as a mere means.

Finally, consider an act consequentialist account, on which we should keep our promises because acts of promise-breaking are bad in themselves. Why think that promise-breaking is bad in itself? Because (we might say) the practice of making and keeping promises is valuable not only

as a means, but also valuable *as an end*, non-instrumentally (Smith 2011). Compare a musical performance: we might value playing Beethoven's Fifth Symphony well *for its own sake*, not just as a means to pleasure or prestige. And if this is so, it seems that we have reason to play the right notes, because playing the wrong notes would lessen the performance's value. Similarly, the consequentialist can argue that people have reason to keep their promises because doing otherwise would lessen the value of the promising practice.[4]

Could we similarly contend that the act of violating a joint decision is bad in itself? Perhaps. Compare the individual case. Granted, it is not so plausible to think that *deliberation* is valuable for its own sake — should we linger on each decision, so as to maximize the amount of deliberation we get to do? But being an autonomous, self-governing agent *is* plausibly valuable for its own sake; and one important component of such self-governance is following through on the decisions one makes in practical deliberation (cf. Bratman 2009). Along similar lines, we should grant that the activity of joint deliberation itself is not valuable for its own sake. But, we might claim, the activity of *governing our lives together* is valuable for its own sake, and one important component of such 'shared governance' is following through on the decisions we make in joint practical deliberation. Thus we might maintain that the act of violating a standing joint decision constitutively undermines the value of our activity of shared governance, and is therefore bad in itself. Thus the most plausible consequentialist account of the bindingness of promises is at least as (and likely more) plausible as an account of the bindingness of joint decisions.

So, whether your favorite account of promises' normative force appeals to Rawls' principle of fairness, a two-tier theory like contractualism or rule consequentialism, Kant's formula of humanity, or the non-instrumental value of the promising practice, a similar account will vindicate the normative force of joint decisions. Without deciding here which of the above

---

[4] The claim is that promise-breaking acts undermine the value of the promising practice *constitutively*, regardless of their downstream causal consequences. Thus this account avoids the false prediction that you can break promises if you know you won't get caught: even undetected promise-breakings undermine the value of the practice they violate.

accounts is best, we can conclude the following: if promises are binding, then I-proposals are too – and likely for the very same reasons.

1.2. Directedness

Promises don't just create obligations *simpliciter*. They create *directed* obligations: obligations owed *to the promisee* in particular. If I promise you that I will φ, then typically I am thereafter obligated *to you* to φ; my failing to φ would wrong *you*. When I promise you that I will read your paper, I become obligated *to you* to do so: my failing to read your paper would wrong *you*. You have a special standing to hold me accountable to my promissory obligation: you can demand that I fulfill my promise and resent me personally if I break it (Gilbert 2004; Darwall 2011).

Can we say the same for I-proposals? When I make an I-proposal to φ, do I then become not only obligated to φ, but obligated *to my addressee(s)*? I think so. When I make an I-proposal to you and you do not object, the resulting joint decision is one that you and I made *together*. Whether or not I abide by this joint decision matters to you in a way it does not matter to anyone else. As the other party to our joint decision, you have a special stake in its fulfillment, and thus have special standing to complain about its violation. So, when we make a joint decision, we each become obligated to the other to do our part in seeing it through.

We can bolster the case for thinking that I-proposals are similar to promises in this respect by focusing on the promisee's special standing to *demand* that the promise is fulfilled (Gilbert 2004; Darwall 2012). This authority is usually given to the promisee alone: third parties may not have standing to demand the promise's fulfillment. Suppose, for example, that I promise my partner Dee Dee that I will wash the dishes tonight. Our neighbor George knows about this promise, and sees the dishes festering in the sink. He demands: "Brendan, wash the dishes!" Is George's demand legitimate? I don't think so. My friend may be perfectly entitled to *advise* me to wash the dishes, but to *demand* that I do so is to overstep his authority. I could rightly respond, "It's none of your business whether I wash the dishes!" (cf. Gilbert 2004: 101). Only *Dee Dee* is in

a position to hold me to my promise: she alone has the authority to demand that I wash the dishes.

I submit that I-proposals have the very same effect: they give their addressees special standing to demand that the speaker abide by the joint decision she proposed. Recall the deliberative theory of demands: a demand is a proposal to make a joint decision regarding the addressee's actions by means of propose-and-challenge.[5] Given this hypothesis, it is easy to see why I-proposals give their addressees standing to demand their fulfillment. A joint decision that the speaker proposes by saying "I will ϕ" (an I-proposal) will be expressed by the addressee by saying "you will ϕ" (a demand). Thus the addressee's demand that the speaker act in accordance with her I-proposal expresses *the same joint decision* that the speaker proposed in the first place. If the initial I-proposal was warranted, then the addressee's demand will be too, since these two speech acts express the very same joint decision.

So, we can explain our example as follows. According to the deliberative theory, my promise to Dee Dee brought into force between us a joint decision to the effect that I will wash the dishes. Dee Dee, as a participant in the process of joint deliberation within which this decision was made, has the standing to express this decision by demanding that I wash the dishes.[6] However, my neighbor George is *not* a participant in the joint deliberation I share with Dee Dee – he and I have our own, separate process of joint deliberation. There is no standing joint decision *between us* to the effect that I will wash the dishes. Moreover, *we* may not be warranted in jointly deciding that I wash the dishes, for the reasons I share *with George* may not sufficiently support my washing the dishes, though the reasons I share *with Dee Dee* do. Whether or not third parties can demand compliance with a promise is thus a contingent matter, depending on the extent to which the reasons they share with the promisor are at stake in the

---

[5] I defend this theory of demands in Chapter 4.

[6] But hasn't the joint decision that I will wash the dishes already been made? How, then, can the demand propose it again? Rather than seeing Dee Dee's demand as proposing the joint decision be drawn anew, I think we should see it as *reiterating* our joint decision. Seeing that I am liable to violate our joint decision, Dee Dee expresses it again. Compare reiterating an assertion: I've told you that it's raining, but you walk out the door without an umbrella. I say again: "it's raining out there!" I am not telling you something new; I am reiterating the assertion I already made.

promise's fulfillment. In contrast, the promisee's standing to demand fulfillment is far less contingent, as it is based directly in the joint decision that the promise brought into force. Thus the deliberative theory predicts that this central feature of promissory obligation – the promisee's special standing to demand the promise's fulfillment – will attach to the obligations created by I-proposals as well.

1.3. Exclusion

Most agree that our reason to keep our promises is only *pro tanto* and can be overridden if enough is at stake. But some have argued that it is still inappropriate for a promisor to think of the fact that she promised to φ as *just another consideration* to be weighed up when deciding whether to φ (Raz 1977; Hart 1982: 255; Robins 1984; Owens 2012: 89-91). Why? Because a promise to φ should *settle the question* of whether to φ for the promisor. To weigh up reasons for and against keeping one's promise is to open a question that one should be treating as closed.

The attitude one should take towards actions one has *promised* to do is tellingly similar to the attitude theorists of intention have argued one should take towards actions one *intends* to do (Bratman 1987; Holton 2004). Like promises, intentions settle deliberative questions, excluding options incompatible with the intended action from further consideration. Moreover, intentions resist reconsideration: having decided to φ, a rational agent should be disposed not to reopen the question of whether to φ unless her situation changes in some unexpected way.

Given the analogy we have pursued between joint and individual deliberation, these observations about individual intentions should apply to joint intentions as well. So, just as I should be disposed to exclude options incompatible with my standing decisions from my deliberation, I should similarly be disposed to exclude options incompatible with our standing *joint* decisions from my deliberation. And just as I should not be over-ready to reconsider my intentions, I should not be over-ready to reconsider our joint decisions. If I fail to have these dispositions, this means that I am not taking our joint decisions as seriously as I take my own individual intentions. Thus, like promises, I-proposals have exclusionary force: a successful I-

proposal that I will φ, by bringing a joint decision into force, will typically make it inappropriate for me to seriously consider the question of whether to φ in my subsequent practical deliberation.

The deliberative theory can thus *explain* the exclusionary force of promises by appeal to the exclusionary force of intentions (cf. Raz 1975: 69-70; Robins 1984). One should not think of a promise as just another consideration to be weighed up in deliberation because, in doing so, one fails to treat the joint decision one's promise brought into force as what it is: a *decision*.

1.4. Uptake

Many have observed that promises require *uptake* from the promisee to be binding (Thomson 1990: 297; Darwall 2011: 268; Owens 2012: 224-226; Roth 2016: 89-92). If I make a promise that you neither hear nor acknowledge, then my promise misfires, failing to obligate me at all. But it is hard to say exactly what kind of uptake promises require. For promises do not seem to require explicit affirmation from both parties in the way agreements, offers, and requests do. Promises seem *unilateral* in a way that offers or agreements are not. On the other hand, the promisee must recognize a promise for it to be binding, and can block a promise's effects by refusing it.

Notably, both of these features characterize I-proposals as well. Any proposed joint decision comes into force only if its addressees recognize it and signal their acceptance. But in the propose-and-challenge method, a proposal need not be *explicitly* accepted to come into effect: all that is required is for the addressee to recognize it and refrain from challenging it in good faith (see Ch. 1, §3.2). Thus I-proposals require the same kind of uptake that promises do: recognition and the absence of objection. And if we take on the deliberative theory, we can explain why promises seem unilateral: because they employ propose-and-challenge.[7]

---

[7] Hallie Liberto (forthcoming) has recently raised several cases in which promissory obligation appears to come into force before uptake is secured. Here's one: on Monday, Bertha sends Albert a letter promising to be monogamous; knowing that Albert won't receive the letter until Friday, Bertha has one last romp with her ex on Thursday. Bertha's action seems wrong, indicating that her promise was binding *before* it received uptake from Albert.

Prima facie, this seems problematic for the deliberative theory. For surely Bertha and Albert can't count as having made a *joint decision* until after Albert receives the letter. So this appears to be a case where a promissory obligation comes into force before any joint decision is made.

This account of uptake helps to assuage one *prima facie* worry about the deliberative theory. Perhaps the most immediate objection to the view is: "But promising doesn't *feel* like making a proposal." True enough: on the ordinary use of 'proposal', a proposal must be explicitly accepted to come into force. Still, I call promises 'proposals' because they *do* require their addressees' acceptance to come into force, even though this acceptance is usually tacit. Here assertion is a helpful analogy. On a popular account, an assertion is "a proposal to change the context by adding [its] content to the information presupposed" (Stalnaker 1999: 10). But assertions, like promises, do not require explicit acceptance to come into effect; they only require their addressees to recognize them and refrain from objecting. So, promises are proposals in the same technical sense that assertions are.

1.5. Release

Suppose I promise my partner Dee Dee that I'll wash the dishes tonight. But before I pick up the sponge, Dee Dee says, "On second thought, don't worry about doing the dishes. You should relax instead." Dee Dee has *released* me from my promise: I am no longer obligated to do the dishes. The moral situation is much like it would have been if I had never promised at all. In general, promisees have the power to release promisors from their promissory obligations.

The addressee of an I-proposal has a similar power: he can propose to retract the joint decision the I-proposal brought into effect. Just as individual decisions can be retracted, so too can joint decisions. When we retract a joint decision, its normative effects are reversed: our reason to abide by our joint decisions no longer gives us reason to abide by *this* decision, as it no

---

True, but Liberto's cases are surprising for *any* theory of promises that includes an uptake requirement. The lesson Liberto draws from these cases is not that promises do not require uptake, but that uptake can have a backward-reaching effect, retroactively making it the case that the promisor was under a promissory obligation from the moment she 'launched' the promise. I see no reason why the uptake of a joint decision cannot have this backward-reaching effect as well. Indeed, Bertha's Thursday night romp seems no less wrong if her letter proposes an explicitly joint decision: "I'm ready. Let's be monogamous." (Curiously, the backward-reaching effect only seems to arise for propose-and-challenge proposals. If Bertha wrote, "I'm ready. Would you like to be monogamous?" her Thursday romp would be less objectionable). So while the backward reach of uptake is a surprising phenomenon in need of explanation, I do not think it poses any special problem for the deliberative theory.

longer stands. So, the addressee of an I-proposal can release the speaker from the obligation it creates by proposing to retract the joint decision the I-proposal brought into force.

But this seems to miss an essential detail. *Any* participant in joint practical deliberation can propose to retract a joint decision, including the person who originally proposed it. The power to retract joint decisions thus seems to be symmetrical. In contrast, the power to release someone from a promise is *asymmetrical*. A promisor cannot release *herself* from her promise; only the promisee can release her. How can the deliberative theory explain this asymmetry?

The answer is that both promisor and promisee *can* propose to retract the joint decision brought into force by a promise, but their proposals should be evaluated by different methods. A promisor's proposal to retract should be evaluated by the *propose-and-ratify* method, while a promisee's proposal to retract should be evaluated by the *propose-and-challenge* method. Why? To correct for both parties' natural biases. The promisor is likely to be overeager to retract the joint decision brought into effect by her promise, since the costs of fulfilling it will fall primarily on her. So, it makes sense to evaluate promisors' retraction proposals conservatively, requiring the promisee's explicit acceptance for them to succeed. In contrast, promisees may be reluctant to retract the joint decisions brought into force by promises, since often they stand to benefit from the promise's fulfillment at little cost to themselves. And so it makes sense to evaluate promisees' retraction proposals using the more liberal propose-and-challenge method.

On the deliberative theory, then, what is unique about the promisee's power of release is not that only she can propose to retract the promise, but that only her proposals should be evaluated using propose-and-challenge. When Dee Dee proposes to retract our joint decision ("don't worry about doing the dishes tonight"), I need not explicitly accept her proposal for the decision to be retracted.[8] But suppose instead I turn to Dee Dee and say, "I'm pretty tired. Would you mind if I left the dishes until tomorrow?" I thereby propose to retract the decision

---

[8] This account does imply that I could *challenge* Dee Dee's attempt to release me from my promise. This might be surprising: we rarely object to being released from our promises. But sometimes we do. Say you promise your proud but frail grandmother that you will help her move into her new home. Being proud, she tries to release you from your promise: "You don't have to help, I'll be fine." Since she *does* need your help, you reply: "I insist: I gave you my word and I'm sticking to it!" You thereby challenge your grandmother's proposal to release you from your promise.

57

brought into force by my earlier promise. But notably, my proposal will only succeed with Dee Dee's explicit blessing. If she says, "Sure, that's fine," our joint decision is then retracted, and I am released from my promise. But if she rejects my proposal ("I'm sorry, but the dishes need to be done tonight"), my promise remains in force. This is just what we should expect if promisors' retraction proposals are evaluated by the propose-and-ratify method (cf. Gilbert 2011b: 99-100).

This asymmetry in deliberative methods guarantees that the speaker of an I-proposal will be released from her resulting obligation *only if* her addressee explicitly agrees to her release: either by proposing to retract the decision himself, or by explicitly accepting the speaker's proposal to do so. No such power is given to the speaker: the addressee does not require the speaker's explicit acceptance to release her. Thus just as promises give promisees the power of release, I-proposals give their addressees a tellingly similar power to reverse their effects.

2. The validity conditions of promises

Sometimes, promises don't bind at all. A person promises to φ, and yet, due to unusual circumstances, her promise fails to create any reason for her to φ (at least in the usual way). Such promises are called *invalid*.[9] The paradigm cases of invalid promises are those elicited by coercion and by deception, as in these colorful examples from Judith Jarvis Thomson:

> **BANK:** "[Bert] is not among the more efficient extortionists: he holds a gun to [Anna]'s head and says 'GIVE ME YOUR WORD THAT YOU WILL GO TO YOUR BANK AND FETCH ME BACK A THOUSAND DOLLARS OR I'LL SHOOT YOU!' [Anna] says 'Yessir'" (Thomson 1990: 310).

> **VAN GOGH:** Diego offers to sell Cora his Van Gogh for $1000, and Cora promises to go to the bank and get $1000 to pay him straightaway. On the way, Cora learns that the painting is forged, and Diego knew this was so (paraphrase of Thomson 1990: 312-313).

---

[9] I assume that invalid promises are still *promises*: a promise was made; it just failed to generate any reasons.

The promise in BANK is invalid due to coercion; the promise in VAN GOGH is invalid due to deception. These background conditions somehow rob the promises of their moral force: neither promisee would be wronged if the promisor did not give him the promised money.

Are I-proposals also invalidated by coercion and deception? I will argue that they are. In particular, I will argue that I-proposals induced by coercion or deception *fail to result in the making of the joint decisions they propose*. There is a gap between the making of an I-proposal and the making of a joint decision: the former can fail to bring about the latter. A simple way for this to happen is for the proposal to be *rejected*: if you refuse my I-proposal, no joint decision is made. This is not what happens in BANK and VAN GOGH: in both cases, the addressee accepts the speaker's proposal wholeheartedly. Yet, I claim, the interaction between speaker and addressee in these cases fails to constitute the making of a joint decision.

Why? Because coerced and deceived promises cannot be accepted in good faith (see Ch. 1, §3.2). Good faith participation in joint practical deliberation requires one to accept only those proposed joint decisions one believes to be warranted by the shared reasons. But, I shall argue, when someone induces a promise via coercion or deception, they are always in a position to know that the promise they have induced is unwarranted. Thus, if she is deliberating in good faith, the recipient of a coerced or deceived promise would challenge that promise on the basis that it is unwarranted. By intentionally withholding that challenge, the promisee accepts a joint decision she knows to be unwarranted, and so cannot be deliberating in good faith.

As I argued in Chapter 1, bad faith joint deliberation is not really joint deliberation at all. The person who deliberates in bad faith is only pretending to engage in joint decision-making with her fellow participants. Thus one person's bad faith renders any resulting 'joint decisions' null and void: her failure to play her part in the deliberative process blocks the kind of joining of wills that is essential to joint decision-making. Since, according to the deliberative theory, the normative force of promises is grounded in the normative force of joint decisions, we can predict that a promise accepted in bad faith will fail to bind the speaker to action. This suggests the following account of promises' validity conditions:

The Good Faith Account of Validity: A promise is valid *if and only if* it is made and accepted in good faith. An accepted promise is invalid when either the promisor made it in bad faith, or the promisee accepted it in bad faith.[10]

I will now argue that the good faith account predicts and explains our nuanced intuitions about when promises are and are not valid. But first, a stylistic note. In section 3, I treated I-proposals as distinct from promises, so that we could investigate their properties without assuming the deliberative theory. In this section, I will abandon this device, taking on the deliberative theory's thesis that promises *are* I-proposals and seeing what follows from this assumption. This will help us to see how the deliberative theory not only captures the validity conditions of promises, but *explains* them.

2.1. Deceived promises

Begin with deceived promises. In VAN GOGH, Cora promises to pay Diego $1000 for his Van Gogh. But she only makes this promise because she believes the painting to be genuine, while Diego knows it to be a forgery. Diego's deception renders Cora's promise invalid. Why?

On the deliberative theory, Cora's promise is a proposal to jointly decide that she will go fetch $1000 in payment for Diego's painting. If the painting were a genuine Van Gogh, this joint decision would be warranted. But Diego knows that the painting is a forgery. So, the reason Cora takes to warrant jointly deciding that she will fetch Diego $1000 – namely, that he will give her a Van Gogh in return – is not a *shared* reason for Diego and Cora, since Diego does not believe it to

---

[10] Here the details of the deliberative theory substantively differ from those of Margaret Gilbert's similar account (2011b). (Recall, however, that the most important difference lies in the different *arguments* Gilbert and I give for the view; see footnote 1). Gilbert also holds that valid promises bring into force a joint decision regarding what the promisor will do; but she gives a different account of the process whereby joint decisions are made. Gilbert holds that a joint decision J comes into effect just in case (a) all parties to the decision have expressed their readiness to be jointly committed to J and (b) this fact is common knowledge between them (1993a: 695; 2011a: 311). Importantly, the parties to the joint decision may express their readiness to accept it *for whatever reason*, including reasons that depend upon their having been coerced or deceived. Thus Gilbert's theory predicts that promises made under coercion or deception result in *bona fide* joint decisions and so are binding (a result she embraces: see her 1993a). In contrast, as I will argue presently, the deliberative theory entails that coerced and deceived promises bring no joint decisions into force and thus are not binding. So, the account of promises' validity conditions I offer here depends upon a very non-Gilbertian theory of joint practical deliberation.

be true. This follows from UNANIMITY (Ch. 1, §1.3), given the plausible assumption that p cannot be a reason for someone who does not believe that p. Since Diego does not believe that the painting is a Van Gogh, this 'fact' cannot be a reason for him individually; and if it is not a reason for Diego, it cannot be a shared reason for Diego and Cora. But without the assumption that the painting is a Van Gogh, Cora's proposed joint decision is not warranted: a forged Van Gogh is not worth $1000 to Cora.

Diego knows all of this. So, if he were deliberating with Cora in good faith, Diego would challenge her promise as follows: "You are only promising to pay me $1000 for this painting because you believe it is a genuine Van Gogh. But I lied to you: the painting is a forgery. Given that the painting is a forgery, you shouldn't pay me $1000 for it, and so I can't accept your promise to do so." Of course, our duplicitous Diego is unlikely to have this sudden change of heart. But that's the point: the fact that Diego does *not* raise this challenge shows that he is not participating in the propose-and-challenge process in good faith, and this is what invalidates Cora's promise. The promise fails to bring its proposed joint decision into force because it was not accepted by means of good faith joint deliberation.

To generalize: say the promisee knows that $\sim p$, but has deceived the promisor into believing that p. When a promise is invalidated by deception, this will be because the warrant of the joint decision it proposes depends upon the truth of p in a certain way: given p, it would be warranted, but without the assumption that p, it is unwarranted.[11] The speaker promises to ϕ because she takes the shared reasons to support her ϕing, given her belief that p. But p is *not* part of the shared reasons, since the promisee believes p to be false (by UNANIMITY). Without p, the shared reasons do *not* sufficiently support the promisor's ϕing, and thus the joint decision that she ϕ is unwarranted (by SUFFICIENT SHARED REASON; Ch. 1, §2). Since he knows that $\sim p$, the promisee must also know that the joint decision proposed by the promise is unwarranted. If he were deliberating in good faith, the promisee would challenge the promise on this basis; the fact

---

[11] Dougherty (2013: 731) makes a similar point about how deception invalidates consent.

that he doesn't shows that he is deliberating in bad faith. Because it is accepted in bad faith, the deceived promise cannot yield a joint decision. This is why deceived promises are invalid.

2.2. Coerced promises

In BANK, Anna's promise is invalid because it is elicited by Bert's coercive threat. Does that mean that *all* promises elicited by threats are invalid? No: consider

> **PICASSO:** Cora wants to purchase a (genuine) Picasso Diego is selling. Cora tells Diego that she is willing to pay $1000 for the painting, but she doesn't have the money on her person. Diego replies: "This Picasso is highly in demand, so I can't hold it for you unless I have a guaranteed sale. Give me your word that you will go to your bank and fetch me back a thousand dollars, or I'll sell the Picasso to someone else." Cora says, "Yessir."

In both PICASSO and BANK, the promisor promises to fetch $1000 only because the promisee threatens to do something if she doesn't. And yet Cora's promise in PICASSO is intuitively valid. So, the mere fact that a promise was elicited *by a threat* is not enough to render it invalid.

The relevant difference between BANK and PICASSO seems to be that Bert's threat to shoot Anna is *illegitimate* – he has no right to shoot her – while Diego's threat to sell the Picasso to someone else is *legitimate* – he has every right to do so (cf. Scanlon 1990: 224 and Chwang 2011: 160-161). What does this distinction amount to? Following Japa Pallikkathayil (2011: 12), I suggest that a threat is illegitimate just in case its addressee can warrantedly *demand* that it not be carried out. So, Bert's threat is illegitimate because Anna can warrantedly demand of Bert that he refrain from shooting her, while Diego's threat is legitimate because Cora cannot warrantedly demand of Diego that he refrain from selling the painting to someone else. Our datum, then, is that promises elicited by *illegitimate* threats are invalid. Why is this the case?

When someone makes a threat, she attempts to unilaterally exclude a certain shared option from consideration. When I threaten, "If you φ, I'll [insert threat here]," I am attempting to rule out the shared option in which you φ and I don't carry out my threat. If my threat is illegitimate, this restriction cannot be warranted. To show this, we must appeal again to the

hypothesis that demands propose joint decisions regarding the addressee's actions. Given this hypothesis, if you can warrantedly demand that I not carry out my threat, the joint decision that I will not carry out my threat must be warranted. Thus my threat restricts our shared options in exactly the wrong way: it attempts to rule out the option in which you φ and I *don't* carry out my threat, while we should in fact rule out the option in which you φ and I *do* carry out my threat. Thus, if you can demand that I not carry out a threat, our shared options should reflect this, letting you do as you please without fearing that I will do what I threatened.

With this in mind, consider BANK. When Bert makes his threat, he unilaterally rules out the shared option in which Anna retains both her money and her life. Anna is left with a choice between promising Bert $1000 and being shot; she understandably opts for the former. But Anna's promise is premised on the wrong set of shared options. Anna has the standing to demand that Bert not shoot her *whether or not* she promises him $1000. Thus her choice should not be between *promising Bert $1000* and *being shot by Bert*, but instead between *promising Bert $1000* and *not promising Bert $1000*, with Bert's shooting her squarely off the table.

Bert is in a position to know all of this. So, if he were deliberating with Anna in good faith, Bert would challenge her promise as follows: "You're only promising to go fetch me $1000 because I threatened to shoot you if you didn't. But I should never have threatened to shoot you in the first place – doing so would be wrong. So let's agree that I won't shoot you, and then you can decide whether or not to promise me $1000 on your own." Of course, our belligerent Bert won't raise this challenge unless he has a sudden and miraculous change of heart. But the fact that Bert *doesn't* raise this challenge, which he knows to be warranted, shows that he is not deliberating with Anna in good faith. So the proposed joint decision that Anna will go fetch Bert $1000 does not come into force, as it was not jointly made. This is why Anna's promise is invalid.

To generalize: promises elicited by illegitimate threats are necessarily premised on a flawed set of shared options. Say you promise to φ because I illegitimately threaten, "if you don't φ, then I'll [insert threat here]." You are only promising to φ because you prefer the shared option *you φ, I don't carry out my threat* over the option *you don't φ, I carry out my threat*. But this is a

63

false choice, imposed on you by my illegitimate threat. Since you would be warranted in demanding that I not carry out my threat, we should rule out the shared option *you don't φ, I carry out my threat* and instead open the option *you don't φ, I don't carry out my threat*. The relevant deliberative question is whether the proposed joint decision is warranted relative to *this* set of shared options. Since I am aware of this, I should challenge your promise on the basis that it is premised on the wrong set of shared options. If I do not raise this challenge, I cannot be deliberating in good faith, and so your promise is invalid.

Crucially, this argument does not extend to promises elicited by legitimate threats. These promises are also premised on a restricted set of shared options. But the fact that the promisor *cannot* demand that the promisee not carry out his threat implies that the promisee has warrant to restrict the shared options in this way. When, in PICASSO, Diego tells Cora, "I won't hold the painting for you unless you promise me payment," he is unilaterally ruling out the shared option in which he holds the painting for Cora without a promised sale. But since Cora cannot demand that Diego do this, the resulting restricted set of shared options is warranted: it simply takes Diego's permissible plans into account. Thus Cora's promise can be warranted despite being premised on Diego's threat, and Diego can accept it in good faith. This is why promises induced by legitimate threats can be valid, though those induced by illegitimate threats are not.

2.3. Comparison with other accounts

I know of two major alternative accounts of the validity conditions of promises:

The Freedom Account: an accepted promise is invalid if and only if the promisor did not make it freely and voluntarily.

The Wronging Account: an accepted promise is invalid if and only if the promisee elicited it by means of wronging the promisor.

How do these alternatives compare to the good faith account?

Begin with the freedom account. This view gets its great *prima facie* plausibility from the fact that the paradigm case of an invalid promise, a coerced promise, is also a paradigm case of

64

unfree action. When I make a promise at gunpoint, I clearly do not make that promise freely and voluntarily, and so it is plausible to suppose that this explains why my promise is not binding. Perhaps this is why many theorists simply assume the freedom account without argument.[12]

Though the freedom account may be able to capture our intuitions regarding coercion, it runs into trouble in explaining invalidity by deception. To explain these cases, the freedom account must appeal to the excuse from responsibility that deception provides. Deception can excuse by showing that it was rational for the deceived agent to do what she did, given what she reasonably believed to be true. For example, we would not criticize Cora for promising to buy a forgery, since she reasonably believed it to be a genuine Van Gogh. Perhaps the fact that Cora's promise can be excused in this way shows that it was not made freely. Accepting this premise would enable the Freedom Account to predict that deceived promises are invalid. But this strategy overgenerates. Deception excuses because the deceived agent is ignorant of some relevant matter of fact. So, if deception by the promisee renders a promise unfree, so too will simple ignorance or deception by a third party. Consider

> **DEGAS:** Cora promises Diego that she will go fetch $1000 to pay for Diego's (genuine) Degas. Cora is buying the Degas as a gift for her friend Ellen, whom she believes is a great lover of Degas' work. In fact, Ellen hates art, but told Cora that she loves Degas in order to impress her. If Cora knew this, she would not have made her promise to Diego.

The promise in this case is clearly valid. If Cora reneged on her promise, she could not justify herself to Diego by explaining that she was deceived about Ellen's preferences. Yet the freedom account cannot plausibly claim that deception by the promisee invalidates promises without also maintaining that deception by third parties invalidates promises. In contrast, the good faith account predicts the difference between these cases. If a promisee elicits a promise by deception, then he can only accept it in bad faith. But if a promise is based in deception by a third party,

---

[12] Rawls, for example, explains the validity conditions of promises by appeal to the requirement that "[the promisor's] words must be spoken freely or voluntarily" (1971: 345). Similarly, Scanlon writes into his account the requirement that "A *voluntarily and intentionally* leads B to expect that A will do X" (Scanlon 1998: 304; my emphasis). For explicit defense of the freedom account, see McMahon (1989: 241) and Chwang (2011).

then so long as the promisee is also ignorant of the relevant fact, he can accept the promise in good faith.

Let us turn to the wronging account, which has recently been defended by David Owens (2007; 2012). Owens argues that promises elicited by wrongful means are invalid because "the promisor's interest in (say) expressing himself by changing the normative situation in favour of the promisee will not be served where the promise has been given only because the promisee has flouted some of the norms governing relations between them" (Owens 2012: 243). On this basis, Owens concludes: "promises induced by duress or misrepresentation are invalid where by getting someone to promise in that way you wrong him" (Owens 2007: 299).

The wronging account and the good faith account yield the same verdicts in nearly every case. Plausibly, deliberating in bad faith wrongs one's co-deliberator; so, whenever the promisee accepts a promise in bad faith, he wrongs the promisor as well. Thus whenever the good faith account predicts a promise to be invalid because the promisee accepted it in bad faith, the wronging account will agree (because the promisee wronged the promisor). However, the accounts diverge in at least one kind of case: when a promisee induces a promise by wronging the promisor, but does *not* deliberate in bad faith. Here I think the good faith account yields the correct verdict. Consider two examples:

> **COMIC BOOKS:** Cora is buying paintings from Diego again. Before they get to business, Cora and Diego are making small talk. Cora notices that Diego has a large collection of superhero comics on his bookshelf, and casually inquires, "I take it you like comic books?" Now, Diego *does* like comic books, but he knows that Cora is an obsessive comic book fan – so much so that, if she gets started talking about comic books, she'll talk so long that she will never get around to buying a painting. In order to avoid this eventuality and thus get Cora to promise to buy a painting, Diego lies: "Oh, those are just for my nephew – I've never read one myself." The conversation moves on to business, and Cora promises to pay Diego $1000 for his (genuine) Rembrandt.

PUSHY ACTIVIST: Ellen and Frank are together in an elevator. Suddenly, Frank presses the emergency stop button. "I work for the Animal Rights Corps," Frank says. "We need donations, but we can never get people to sit still long enough to hear our case. I have the key to this elevator, and I'm not letting you out until you listen to what I have to say." By these means, Frank intends to get Ellen to promise a donation to his organization. Ellen begrudgingly listens to Frank's spiel. But as she hears his arguments, Ellen is convinced of the importance of his cause. When Frank is done, he unlocks the elevator. He says, "You're free to go now, but if you feel compelled by what I said, please pledge a donation." Ellen replies, "Though it was wrong of you to lock me in here, I was convinced by your arguments. I promise to donate $1000 to your organization."

It seems to me that Cora and Ellen's promises in these cases are both valid. Cora's promise to pay Diego for the Rembrandt seems binding despite Diego's fib. If Cora were to find out on the way to the bank that Diego lied to her about his comic books in order to prevent her from talking his ear off, she might be understandably chagrined, but could not reasonably conclude that she was off the hook. Diego still has a legitimate claim to payment. Similarly, Ellen's promise of a donation to Franks's organization seems binding despite the objectionable way in which Frank solicited it. Though Frank forced Ellen to listen to his arguments, he did not force her to promise the donation. So Ellen should hold to her word, even if she remains understandably angry with Frank.

But in both of these examples, the promisee elicits the promise *by means of* wronging the promisor. Diego wronged Cora by lying – but if he had not done so, she would have talked his ear off and never gotten around to promising to buy his painting. Frank wronged Ellen by locking her in the elevator – but if he had not done so, she would not have listened to his arguments and been convinced to promise a donation to his organization. Thus both of these cases are examples of the promisee getting the promisor to make a promise by means of an action that wrongs her. So, the wronging account predicts that the resulting promises are invalid.

It might be objected that the causal link between the wronging and the promise in these cases is too indirect for the wronging account to apply. As I granted, Frank did not force Ellen *to promise a donation*, just to listen to his spiel; and Diego did not deceive Cora *into promising*, just into not talking his ear off. Thus the promisee's wrongdoing did not bring about *the promise itself*, just the situation that enabled the promise to be made. The challenge is to find a non-*ad-hoc* rationale for saying that this distinction is relevant to the validity of promises. If it is in our interest to count promises induced directly by wronging as invalid, why wouldn't it also be in our interest to count promises induced *indirectly* by wronging as invalid? The advocate of the wronging account must say more if she is to explain why these promises, brought about by wrongful means, are yet valid.

In contrast, the good faith account provides a principled basis for distinguishing these cases from the paradigm cases of coerced and deceived promises. Though Diego and Frank elicited their promises by means of deception and coercion, they did not accept those promises in bad faith. The fact that Diego enjoys comic books is irrelevant to the warrant of Cora's promise to buy the Rembrandt. So, Diego can correctly judge Cora's promise to be warranted, and accept it in good faith. Since Ellen's options are no longer constrained by Frank when she makes her promise, it is not based in a flawed conception of the options in the same way that Anna's promise in BANK is. So there is no challenge that Frank ought to raise to Ellen's promise – no threat that he ought to withdraw – and he can thus accept it in good faith. While the wronging account would have to contort itself to accommodate these cases, the good faith account explains them straightforwardly.

2.4. Insincere promises

While coercion and deception involve the *promisee* deliberating in bad faith, the good faith account also predicts that a promise can be invalidated by the *promisor's* bad faith. This has a surprising, and admittedly counterintuitive, implication: insincere promises – i.e., promises that

68

one does not intend to keep – are invalid.[13] For, unless she is akratic, a promisor who does not intend to keep her promise will almost always judge that she does not have sufficient reason to do so. Thus she cannot take her promise to be warranted (by JOINT-INDIVIDUAL COHERENCE), and so cannot make it in good faith. Insincere promises involve the same absence of mutuality that invalidates coerced and deceived promises; so, the good faith account implies that they are invalid as well.

This prediction is problematic, since insincere promises intuitively seem to be binding. Suppose I promise to come to your talk while having no intention of doing so. Still, it seems I am obligated to keep my promise – when I don't show up, you should feel wronged. And if you confront me for breaking my promise, it would be absurd for me to defend myself by saying, "Oh, that promise wasn't binding, since I didn't intend to keep it."

However, I stand by the deliberative theory's prediction that insincere promises are invalid. To defend this claim, I will offer an error theory, showing how our intuitions might be explained under the assumption that insincere promises are not binding. Then, I will offer an independent argument for the view that insincere promises are not valid.

My error theory begins from the observation that making an insincere promise is itself wrong. Say that, after insincerely promising to do so, I come to your talk anyway. If you then find out that my promise was insincere, you still have standing to complain. By making a promise I didn't intend to keep, I wronged you, even though I ultimately kept it. I want to suggest that this is also why you should feel wronged when I *don't* show up to your talk: not because I did not do what I promised, but because I promised what I did not intend to do.

Moreover, even if insincere promises are not binding, there are other reasons why an insincere promisor will typically be obligated to do what she promised anyway. First, my insincere promise led you to believe that I will be at your talk. You may have then relied on this

---

[13] Neither the freedom account nor the wronging account considered in the previous section predict that insincere promises are invalid. *Prima facie*, this is a point in these other accounts' favor. Whether the point is decisive will depend on how convincing you find the arguments I offer in the remainder of this section.

expectation in ways that would cost you if it were now disappointed. The best way for me to avoid inflicting these costs on you is to do what I promised, so that your expectation turns out to be true after all. Second, since I wronged you by promising insincerely, I am obligated to make amends for my wrongdoing. Typically, actually *doing* what I insincerely promised to do will be the best first step towards making amends. Keeping my promise will not only prevent you from suffering further harm, but will also demonstrate my commitment to being more trustworthy in the future. Thus, even if my insincere promise does not bind in the normal way, I will still usually be obligated to keep it.

Why, however, should we accept this roundabout account of the obligation to keep insincere promises, instead of the much simpler explanation that these promises are valid? Because, as I shall now argue, there is independent reason to think that sincerity is a validity condition for promises.

Promises communicate intentions: by promising to φ, I communicate that I intend to φ. This is why promising without intending is *insincere*: it communicates that one has an intention one does not have. Indeed, to say "I promise to φ, but I don't intend to φ" seems to be self-undermining in the same way that the Moorean assertion "p, but I don't believe it" is. But if intending to do what one promises is not a validity condition for promises – that is, if insincere promises can be valid – then this observation becomes puzzling. If I can make a perfectly valid promise to φ without intending to φ, then why can't I do so openly, promising to φ while declaring that I have no intention to do so?

We could answer this question easily if we adopted the expectation theory of promising (see §4.2 below). On this view, to promise to φ is to lead one's promisee to form a particularly strong expectation that one will φ. Thus saying "I promise to φ, but I don't intend to φ" will be self-undermining because declaring one's lack of intention undermines the promisee's expectation that one will φ. However, if we reject the expectation theory – which, as we shall see,

70

there is good reason to do (§4.2) – then we cannot avail ourselves of this answer.[14] If we think that promises are distinct from testimony regarding one's future conduct, then the fact that promises communicate intentions requires a different explanation.

Perhaps we can argue as follows: when I promise to do something, I become obligated to do it; and when I am obligated to do something, I should intend to fulfill that obligation. So, a promise to ϕ will implicate that one intends to ϕ, since the latter ought to accompany the former. But this understates the connection between promising and intending. For obligation in general, the implicature can be cancelled: one can undertake an obligation while publicly declaring an intention to violate it. Say I step on your foot, thereby incurring an obligation to step off. I might do so while saying: "I know I am obligated to get off your foot, but I have no intention of doing so. So *there!*" Though spiteful and perhaps akratic, my behavior is not unintelligible.[15] But I think it *would* be unintelligible for me to say: "I promise to come to your talk, but I have no intention of doing so. So *there!*" This promise would not even accomplish my spiteful goal of creating an obligation to violate – it seems to misfire from the outset. Thus the connection between promises and intentions seems to be even tighter than the connection between obligations and intentions.

The point is even clearer if we move to the weaker claim that the effect of a promise is to give oneself *pro tanto* reason to do what one promised (§1.1). For it is not even akratic to give oneself *pro tanto* reason to ϕ while openly intending to do the opposite: I might make it more difficult for myself to finish the race, say by refusing water, in order to demonstrate the strength of my commitment to doing so. Why, then, can't I coherently *promise* to not finish the race in order to demonstrate the strength of my commitment to finishing it? If intention is not a validity condition for promises, then such a promise would be binding. Why, then, would it be so strange to make it?

---

[14] Moreover, our error theory explains the obligation to keep insincere promises by appeal to the very considerations the expectation theorist takes to explain promissory obligation in general (namely, the promisee's expectations). Thus the expectation theory attributes no more force to insincere promises than the deliberative theory does.

[15] Sensitive to this very point, Owens (2008) concludes that promises do not communicate intentions at all. He argues that there is nothing self-undermining about promising to do something while denying that you intend to do it. In this case, Owens' *modus ponens* is my *modus tollens*: I find it far less strange to accept that insincere promises are invalid than to accept that "I promise to ϕ, though I don't intend to" is perfectly coherent.

If, on the other hand, we accept that intending to do what one promises is a necessary condition for one's promise to be valid, we get a simple account of why promises communicate intentions. In general, any speech act presupposes that the conditions of its own validity are met. When I command you to do something, I presuppose that you are able to do it, since my command cannot be binding if it is impossible to obey. When I testify that p, I presuppose that I know p is the case, since I cannot give you knowledge I do not have. This is why it is insincere to command you to do something I know to be impossible, or to testify something I don't take myself to know. This is also why it is self-defeating to say "I command you to ϕ, though you can't" or "p, though I don't know it." These statements purport to do something while, in the same breath, denying the possibility of its being done. Similarly, if intending to ϕ is a necessary condition for validly promising to ϕ, then in purporting to make a valid promise, I presuppose that this condition is met. This explains why promises communicate intentions. It also explains why it is self-undermining to say, "I promise to ϕ, though I don't intend to ϕ": because the truth of the second clause is incompatible with the performative success of the first.

Thus I suggest that the best explanation of why promises communicate intentions is that one must intend to do something in order to validly promise to do it. And this is because one cannot promise in good faith to do something one does not intend to do. Despite appearances to the contrary, then, perhaps we should conclude that insincere promises are not valid after all.

3. An objection: are joint decisions weaker than promises?

Suppose the arguments so far have convinced you that promises and joint decisions have a shared pattern of normative effects. Still, you might have a nagging feeling that promises are somehow more *forceful* than joint decisions. To make a promise is to undertake a solemn moral obligation, which would be a serious wrong to break. But making a joint decision seems a less serious matter. To make this objection vivid, consider two examples:

PLANNED WALK: On a Saturday morning, I propose to my partner Dee Dee that we take a walk in the afternoon. She agrees. Later, I tell Dee Dee, "Sorry, I can't go walking after all – I need to finish writing this section."

PROMISED WALK: I propose a Saturday afternoon walk to Dee Dee; she agrees. Then, remembering my tendency to abandon our walking plans, Dee Dee says: "Okay, but I want to know that you won't change your mind this time. Do you promise you'll go on a walk with me this afternoon?" I say, "Yes, I promise."

Granting that joint decisions generate obligations, the obligations Dee Dee and I undertake to one another in PLANNED WALK still appear fairly weak. If either of us simply forgot to make time for the walk, it wouldn't be a big deal. It also appears to be rather easy to get out of our obligations: telling Dee Dee that I'm busy writing seems to be enough to get me off the hook. In contrast, it *would* be a big deal if, after I promise to go on a walk with Dee Dee in PROMISED WALK, I failed to do so. And I can't get out of this obligation simply by informing Dee Dee that I need to do my work instead; I am bound unless she explicitly releases me from my promise. So, my obligation in PROMISED WALK is stronger than my obligation in PLANNED WALK in at least two ways: it is more important that I fulfill it, and it is more difficult to extract myself from it.

Prima facie, this contrast appears flatly incompatible with the deliberative theory. The deliberative theory seems to predict that the obligations in PLANNED WALK and PROMISED WALK will be identical. After all, it holds that there is *no difference* between my promising Dee Dee that I will go on a walk with her and my proposing to jointly decide that I will do so. Worse, the deliberative theory doesn't seem able to even make sense of the interaction in PROMISED WALK. Dee Dee asks me to promise to do something that we've *already* jointly decided that I will do. On the deliberative theory, this request seems redundant, perhaps even incoherent. But if promising is different from proposing a joint decision, Dee Dee's request makes perfect sense. Thus it seems

the only way to explain the contrast between our examples is to accept that promises are not proposals in joint deliberation.[16]

Not so fast! Appearances notwithstanding, the deliberative theory has the resources to explain the contrast between PLANNED WALK and PROMISED WALK. The trick is to recognize that some joint decisions generate stronger and stickier obligations than others. What makes the obligation in PROMISED WALK stronger than the obligation in PLANNED WALK, I claim, is that the joint decision generated by my promise in PROMISED WALK is *more resolute* and *harder to retract* than the joint decision in PLANNED WALK. Let me elaborate on these two differences.

Begin with resolve. Consider the difference between casually planning to make spaghetti for dinner and solemnly deciding (say, as a New Year's resolution) to go to the gym three times every week. The latter decision is more *resolute* than the former: one is more committed to seeing it through. I might change my dinner plans on a whim; but I will be reluctant to give up my resolution to go to the gym, and disappointed in myself if I do. Similarly, joint decisions can be more or less resolute. Some joint decisions, like a plan to go for a walk this afternoon, are fairly casual; others, like the decision to get married, are more serious. The more resolute our joint decision, the more reluctant we will be to reconsider it, and the more disappointed we will be in ourselves (and each other) if we fail to see it through.

To be clear, resolve thus understood is a descriptive concept. It is a feature of the attitude we take towards our (joint or individual) decisions. But it has a direct normative upshot: plausibly, more resolute joint decisions generate stronger obligations.

A second way in which joint decisions can differ is in their *retraction conditions*. As I argued in §2.4, the joint decisions brought into force by promises have *asymmetric* retraction conditions. While the promisee can propose to retract the joint decision using propose-and-challenge, the promisor can only propose retraction by means of propose-and-ratify. By design, this asymmetry of methods makes it more difficult for the promisor to get out of her obligation, but relatively

---

[16] Thanks to an anonymous reviewer for bringing this objection to my attention, and to Thomas Hurka for helping to sharpen it considerably.

easy for the promisee to release her. In contrast, most agreements (i.e., joint decisions that concern both parties' actions) will have symmetrical retraction conditions. A casual agreement might allow both of us to propose retraction using propose-and-challenge; a more serious agreement might require us to retract by the more explicit means of propose-and-ratify.

Here, then, is how I suggest we understand the contrast between our cases. In PLANNED WALK, the joint decision Dee Dee and I make to go on a walk is weak in resolve and easy to retract. Since our decision is not very resolute, it entails a relatively weak obligation to see it through – which is why it is not a big deal if we fail to do so. And since there's no strong reason to avoid reconsidering our plan, it is plausible that both Dee Dee and I can propose to retract our joint decision using propose-and-challenge. This is exactly what happens when I say "Sorry, I can't go walking after all" – and since Dee Dee does not object, our plan is thereby withdrawn.

In PROMISED WALK, our initial joint decision (before Dee Dee asks me to promise) is just like the decision in PLANNED WALK. But Dee Dee realizes that so long as our plan is casual and easy to retract, I am likely to abandon it. So she asks me to promise. My promise brings into force a joint decision that has the same content as our prior decision (that we will go on a walk), but that is more resolute and harder to retract. Since the resulting decision is more resolute, my obligation to go on a walk with Dee Dee is stronger, and can't be overridden by (e.g.) my desire to finish some writing task. And our joint decision now has the asymmetric retraction conditions characteristic of promises. I now need Dee Dee's explicit permission to be released from my obligation to go on a walk; I can't get off the hook just by saying "Sorry, I can't go." When Dee Dee asks me to promise, she is really asking me to re-propose our joint decision with greater resolve and asymmetric retraction conditions. By endowing our joint decision with these features, my promise generates an obligation that is stronger and stickier than the one I had before.

But this raises a further question. Why does Dee Dee communicate her wish to make our joint decision more resolute and harder to retract by asking me to *promise*? She seems to be presupposing that a joint decision made via a promise will be more resolute and harder to retract

than one made via an agreement. What warrants this presupposition? The answer lies in the characteristic functions that promises and agreements play in our everyday lives.

Starting again from the individual case, we can distinguish two functional roles of decisions: a *planning function* and a *committing function*. The planning function of decisions is the role they play in helping us to coordinate our actions over time, in particular by enabling us to treat our prior decisions as fixed when deliberating about what to do (Bratman 1987). For example, having decided to cook spaghetti for dinner makes it easier for me to decide what to buy at the grocery store. Decisions can serve this function without being particularly resolute, so long as they settle the question. The committing function of decisions is the role they play in strengthening our commitment to a course of action, often in order to preemptively ward off temptation to do otherwise (Holton 2004). My decision to go to the gym three times a week serves this function. To successfully play this role, a decision needs to be more resolute and resistant to reconsideration.[17]

Similarly, joint decisions can serve both planning and committing functions. Sometimes the main point of a joint decision is simply to coordinate our plans, as when we decide where to meet for coffee. But sometimes the purpose of a joint decision is to strengthen our commitment to a course of action, as when we decide to train for a marathon together. As with individual decisions, joint decisions that primarily serve a planning function can be low in resolve and easy to retract, while joint decisions that serve a committing function must be more resolute and resistant to retraction.

Here's my hypothesis: while agreements are often used to make joint decisions that primarily serve a planning function, promises are almost always used to make joint decisions that also serve a committing function.

Since agreements concern both agents' actions, they often serve to help us plan and coordinate joint actions. So, for example, our agreeing to meet at Starbucks at 3pm enables each

---

[17] The planning and committing functions are not mutually exclusive, and plausibly all decisions serve both functions to a certain extent. But decisions can differ in the extent to which each function is essential to their purpose.

of us to plan on the assumption that the other will be at Starbucks at 3pm. This planning function is immensely useful in making joint action possible; but a joint decision need not be high in resolve or resistant to retraction to serve this function. Thus it is common to make agreements that are low in resolve and easy to retract, such as the agreement in PLANNED WALK.

Since promises concern only the speaker's actions, they don't automatically earn their keep by enabling joint action. Indeed, if all you need is to be able to plan on the assumption that I will ϕ, then often my simply informing you that I intend to ϕ will be enough. When might we need to go beyond this and jointly decide that I will ϕ? The clear answer is: when I need to give you a stronger commitment than a statement of intention can provide. Thus the contexts in which I have reason to promise will usually be contexts in which I have reason to strengthen my commitment to a course of action. This committing function will be best served by a joint decision with high resolve (so as to strengthen my commitment to the action) and asymmetric retraction conditions (so as to make it difficult for me to back out). And so we can predict that promises will usually result in joint decisions that are high in resolve and difficult to retract.

When Dee Dee asks me to promise, she is appealing to our common knowledge that promises are generally used to make joint decisions that are resolute and hard to retract. We both implicitly understand that, by promising Dee Dee that I'll go on a walk with her, I am proposing a joint decision with these properties. This change in the resolve and retraction conditions of our joint decision, in turn, explains why my obligation in PROMISED WALK is stronger and more difficult to get out of than that in PLANNED WALK.

Zooming out from these examples, the worry with which we began was that promising is a more serious matter than joint decision-making. We can now see that this objection stems from a misreading of a nearby truth. Which is: joint decisions concerning one person's actions are generally a more serious matter than joint decisions concerning both parties' actions. This is because joint decisions concerning both parties' actions often serve a purely planning function, while joint decisions concerning one person's actions usually serve a committing function. Since the paradigm joint decision concerns all parties' actions, it is natural to think that all joint

77

decisions are as casual as a plan to go on an afternoon walk. But this is too narrow a view of joint deliberation. When we fully appreciate joint deliberation's versatility, we see that it can encompass both the weak and revisable obligations of everyday plans and the solemn and stubborn obligation of a promise.

4. Other theories of promising

In this section I situate the deliberative theory in relation to the three most prominent theories of promising in the literature: conventionalism, the expectation theory, and the normative powers theory.

4.1. Conventionalism

Conventionalism is the view that promising is a move in a conventional social practice. Following Rawls, we can understand a *practice* as "any form of activity specified by a system of rules which defines offices, roles, moves, penalties, defenses, and so on, and which gives the activity its structure" (1955: 3). Examples include baseball, voting, weddings, jury trials, and queuing in a line. The conventionalist locates promising within a 'promising practice' that has two roles (promisor and promisee), two moves (promising and releasing), and one rule: if a promisor promises a promisee that she will φ, then the promisor must φ unless (i) the promisee releases her from the promise or (ii) certain invalidating conditions (e.g. coercion or deception) obtain (Jones 1966: 288; Rawls 1971: 344-345; Kolodny and Wallace 2003: 120). This allows the conventionalist to ground the obligation to keep promises in a wider obligation to abide by the rules of beneficial practices (Rawls 1971: 342-50).[18]

---

[18] A clarification. *Speaking a language* is a social practice; so, if we assume (likely falsely) that promises can only be made using language, then there is a boring sense in which promises are moves within a practice. Friends and enemies of conventionalism alike should agree that this uninteresting claim is not its thesis (see Raz 1972: 100; Scanlon 1990: 213-216). One way to avoid this confusion is to distinguish between the claim that promising is necessarily performed *by means of* a move within a practice and the claim that promising *is* a move within a practice. Consider the difference between *hitting a home run* and *making the crowd go wild. Hitting a home run* is a move within the practice of baseball; *making the crowd go wild* is not itself a move within baseball, but instead something you can do *by means of* a move within baseball (cf. Austin 1962 on the distinction between illocutionary and perlocutionary acts). The boring claim is that promising is typically performed *by means of* a move within a practice (uttering a sentence in a language);

After a long reign as orthodoxy,[19] conventionalism has recently fallen out of favor, largely due to the objections raised by T. M. Scanlon in his agenda-setting "Promises and Practices" (1990).[20] This presents a worry for the deliberative theory. For, like conventionalism, the deliberative theory also holds that promising is a move within a social practice: the practice of joint practical deliberation. So we must consider whether the objections that toppled conventionalism also provide grounds for rejecting the deliberative theory. I will discuss the two objections that have been the most influential.

The first objection claims that conventionalism cannot explain why promises generate obligations that are owed *to the promisee* in particular (Scanlon 1990: 221; Darwall 2011: 263-264; Owens 2012: 135). For, the argument goes, conventionalism says that promise-breaking is wrong because it violates the rules of a certain practice. But breaking the rules of a practice plausibly wrongs *everyone* who participates in that practice (excepting oneself). So, if the relevant practice is that of promising generally, then breaking a promise seems to wrong *everybody* who has ever made a promise, not just the promisee.

This argument hinges on the assumption that there is a single promising practice in which all promising agents participate. The conventionalist should reject this assumption. Consider the practice of playing chess. Each pair of chess players has their own distinct game of chess; they are not all playing a single enormous game. So, if I cheat in our game of chess, I *do* wrong everyone else who participates in this practice − but since you are the only other player in our game, this means that I wrong you and you alone. To assume that breaking the rules of chess must wrong every chess-player that has ever lived is to conflate *types* of practices with *token* instances of a practice type. Every chess player participates in the type of practice, chess, in which we are engaged; but they do not all participate in our token game of chess.

the conventionalist's claim is that promising *is* a move within a practice. But if you are skeptical of this distinction, there is a brute force solution: we can say that conventionalism's thesis is that promising is necessarily a move within a practice *other than* the practice of speaking a language. Thanks to Judith Jarvis Thomson for discussion here.

[19] Influentially defended by David Hume (1739/1978: 516-525), conventionalism was popular through the late 20th century, winning the allegiance of John Rawls (1955; 1971), John Searle (1969), and G. E. M. Anscombe (1978), among others.

[20] Some exceptions to this trend are Deigh (2002), Shockley (2007), and Taylor (2013).

The conventionalist should say that promising, like chess, is (usually) a two-person game. (If a promise has two addressees, it is a move within a three-person practice, and so on). When I break my promise to you, I am violating the norms of our two-person practice of making promises to one another. My violation *does* wrong everyone else who participates in this practice – but since you are the only other participant in our token promising practice, this means that I wrong you and you alone. The deliberative theory can co-opt this response: breaking my promise wrongs you in particular because it violates the norms of our two-person practice of joint practical deliberation.

The second objection charges that conventionalism makes the existence and nature of promising too *contingent*. Conventionalism entails that our ability to make promises depends upon the prior existence of a social practice of promising. Since we could conceivably fail to have a practice of promising, conventionalism seems to imply that we could lack the ability to make promises to one another. Yet many have thought that morally competent persons *necessarily* have the power to make binding promises (Locke 1689/1980: 13; Raz 1977: 214-215; Scanlon 1990: 201; Thomson 1990: 303-304; Shiffrin 2008; Pink 2009: 400-401).

Scanlon makes this point vivid with a 'state of nature' case. Suppose that I encounter a stranger on a desert island, standing on the opposite bank of a river. Each of us has lost his weapon on the other's side. Gesturing, I convince the stranger that I will throw his boomerang to him if he first throws my spear back to me. The stranger throws me the spear – and I walk away, leaving the boomerang where it was. Scanlon judges that "what I have done in this example is no less wrong than it would have been if I had promised the stranger that I would return his boomerang ... it is the same kind of wrong" (1990: 201). But the stranger and I are members of societies that have made no contact, and so presumably have no shared practice of promising.

Pressing the objection further, Seana Shiffrin argues that since "the power to promise is an essential tool to forestall and neutralize morally problematic inequalities between people" (2008: 485), we need promises for our relationships to meet minimal conditions of moral decency. If, as is plausible to assume, morally competent persons are necessarily able to form

morally decent relationships, then they must necessarily have the power to promise. This gives us reason to reject the conventionalist idea that "promises are inventions that we could have failed to invent and still gotten by morally, although perhaps less well and less efficiently" (498).

The existence of promises thus seems to be less contingent than conventionalism implies. But the conventionalist is only vulnerable to this objection if she bundles together two claims that can and should be separated: the claim that promising is a move within a social practice, and the claim that this practice is *optional*, in the sense that social interaction would be possible without it. It is easy to run these claims together, because it is natural to assume that *all* social practices must be optional.[21] But this assumption is simply false: some social practices are non-optional.

Consider *assertion*. Assertion has all the signature features of a move in a social practice. It is governed by a rule that dictates which assertions are permissible – the much-discussed 'norm of assertion' (Williamson 2000). Beyond this norm, assertions are governed by many other 'maxims' determining when they are relevant or cooperative (Grice 1989). And, as in all practices, what assertions are permissible in a conversation depends upon what other 'moves' have already been made (Lewis 1979). No wonder philosophers studying assertion, just like those studying promises, have felt the irresistible pull of the analogy with baseball.[22]

Yet few are tempted to say that the assertion practice is socially optional. The ability to share information in the way assertion makes possible seems to be a precondition for cooperative society. And the norms governing assertion are widely thought to be "derived from general principles of rationality, cooperation and/or cognition" (Korta & Perry 2015). Thus though

---

[21] While critics usually attribute both of these claims to the conventionalist, some conventionalists are careful to keep them apart. In fact, Hume, the canonical conventionalist, explicitly denies that the promising practice is optional: "Where an invention is obvious and absolutely necessary, it may as properly be said to be natural as any thing that proceeds immediately from original principles ... Tho' the rules of justice be artificial, they are not arbitrary. Nor is the expression improper to call them laws of nature; if by natural we understand what is common to any species, or even if we confine it to mean what is inseparable from the species" (1739/1978: 484 [3.2.1]). Thus it may be that Hume was never vulnerable to Scanlon and Shiffrin's contingency objection in the first place. Thanks to Jed Lewinsohn for drawing my attention to this point, and to the relevant passage of Hume.

[22] Compare Rawls (1955) with Lewis (1979b). See also Williamson (2000): "someone who knowingly asserts a falsehood has thereby broken a rule of assertion, much as if he had broken a rule of a game; he has cheated" (238). Others who endorse the assertion-game analogy include Wittgenstein (1953/2009), Searle (1969), Brandom (1983), Dummett (1993), and Rescorla (2009); for criticism of this analogy, see Cappelen (2011) and Maitra (2011).

assertion is a move in a practice, this practice is not *conventional* in the Lewisian sense of the term: it is not one of *multiple different* possible solutions to a coordination problem (Lewis 1969/2002: 70). The practice of assertion, with just the rules it has, is the unique solution to the problem of sharing information between persons.

Taking assertion as our model, then, we can make room for a *nonconventionalist practice theory* of promising: a theory that claims, with the conventionalist, that promising is a move in a practice, but departs from conventionalism in claiming that this practice is not socially optional.[23] The deliberative theory occupies this relatively uncharted area of logical space. Like conventionalism, the deliberative theory is a practice theory: it claims that promises are a move within a practice. But unlike conventionalism, the deliberative theory does not take the power to promise to be "activated only by the happenstance that a social convention of promising has developed" (Shiffrin 2008: 483). For the practice within which it locates promises – joint practical deliberation – is plausibly as essential to social life as is promising itself. Like the practice of making assertions, the practice of joint practical deliberation serves a function essential to social interaction: it enables us to plan and act together on terms that all can accept. So it is not a contingent fact that morally competent persons engage in joint practical deliberation; and thus it is no more contingent that we have the power to promise.[24]

---

[23] We could equally well retain the label 'conventionalism' for all practice theories, including the deliberative theory, and simply point out that conventionalism need not imply optionality. I avoid this labeling only because it is potentially misleading, given how widely conventionalism is thought to entail that promises are optional.

[24] Anyone who asserts the necessity of promising has to contend with Korn and Decktor Korn's (1983) argument that there exists at least one society with no practice of promising, in the Tonga Islands of the South Pacific. In Tonga, they claim, statements of the form "I will φ" are not taken to commit the speaker to φing, but instead merely express solidarity for the addressee's desire that φ be done. No alternative form of speech plays the role of promising.

Though this does seem to show that *promising* is not culturally universal, it does not show the same for joint practical deliberation. For it may be that Tongans engage in joint deliberation, but refrain from proposing joint decisions in the particular way involved in promising. Tongans do appear to make demands and requests of one another; they also marry and exchange gifts (Evans 1996). If I am right that demands, requests, and agreements are also moves in joint practical deliberation, this indicates that the Tongans engage in this practice. But why, then, don't they promise? Perhaps because there is a taboo in Tonga against long-term planning, which restricts joint decision-making to the here and now. While it makes sense for me to demand or request that you do something right now, there's no point in promising to do something right now, since I can just *do it*. Promises require a longer-term view. Perhaps this is why Tongans avoid promises while still making joint decisions in other ways.

With this in mind, return to Scanlon's desert island case. The crucial premise of Scanlon's objection is that, since the island stranger and I come from mutually isolated societies, we must share no social practices. But we can now see that this premise is mistaken. To participate in any society at all, one must be competent in those practices that are necessary to social cooperation – including, I have suggested, the practice of joint practical deliberation. Thus the stranger and I are able to deliberate together simply because both of us are members of some society or other. So, I can propose to the stranger that we jointly decide that I will throw him his boomerang if he throws me my spear. On the deliberative theory, to do this *is* to promise him that I will do so. If I walk away without returning the boomerang, then, I do commit "the same kind of wrong" as that of breaking a promise – because I *am* breaking a promise.

4.2. The expectation theory

As an alternative to conventionalism, Scanlon proposed what has come to be known as the *expectation theory* of promising (1990; 1998: Ch. 7).[25] On this view, the obligation to keep our promises derives from a practice-independent obligation to not create false expectations in others. When I promise you that I will ϕ, I give you *assurance* that I will ϕ, leading you to form an strong expectation that I will do so; having deliberatively raised this expectation, it would be wrong for me to violate it.

The expectation theory has substantial intuitive appeal. One of the main reasons why we make promises at all is to give others assurances that they can rely upon. If you are worried about whether I will actually show up to our meeting, I can reassure you by promising you that I will. The fact that I promised to come to our meeting justifies you in having a particularly strong expectation that I will do so, which you can then rely upon in planning your future conduct.

However, the expectation theory doesn't just say that one important function of promises is to give assurance; it says that *all it is* to promise is to give someone assurance. And as the

---

[25] Scanlon was not the first to propose the expectation theory (see Árdal 1968, Narveson 1971, MacCormick 1972, and Thomson 1990: Ch. 12). But Scanlon's is widely regarded as the canonical presentation of the view.

substantial literature in response to Scanlon's proposal has revealed, there is reason to doubt this strong claim. The obligations one undertakes by deliberately raising expectations in others seem importantly different from the obligations one undertakes in making a promise. If I inform you of my future plans to φ, but then change my mind, I can discharge my obligations to you by simply warning you that I won't φ, and perhaps compensating you for any losses you incurred by relying on my testimony. In contrast, if I promise you that I'll φ, merely warning you that I've changed my mind and compensating your losses is not enough: I am obligated to φ, pure and simple (Owens 2006: 53-56; Southwood and Friedrich 2009: 266-271). This difference does not seem to be accounted for by supposing that promises give rise to especially strong expectations. For I can tell you that I will φ with as much confidence as I like, and yet still coherently add the disclaimer, "but I don't promise to φ" (Raz 1972: 99; 1977: 216; Robins 1984: 9; Owens 2006: 60; Darwall 2011: 267). In such a case, it seems that my disclaimer succeeds: I have no promissory obligation to φ. This seems to show that I can knowingly and voluntarily lead you to form a very strong expectation that I will φ without thereby promising to do so.

Scanlon was aware of these worries from the outset (see his 1998: 301-302). His response is to claim that, while not all deliberately created expectations generate promise-like obligations, one undertakes a promissory obligation when one creates a special kind of expectation under a special set of conditions – the conditions encapsulated in his "Principle F" (1998: 304). Even if it succeeds at dodging the above objections, however, this move comes at a steep theoretical cost. The appeal of the expectation theory comes from its potential to demystify promissory obligation by explaining it in terms of a moral obligation that we independently understand: the obligation not to mislead others. By separating promises out from other ways of leading others to form expectations about our actions, and giving promises their own tailor-made moral principle, Scanlon's version of the expectation theory loses this explanatory appeal.

But I won't press this complaint further; my aim here isn't to refute the expectation theory. Instead, I want to show how the deliberative theory can capture both the observation that

motivates the expectation theory – promises' role of providing assurance – and the observation that has motivated some to reject it – the distinction between promises and testimony.

Begin with the assurance-providing role of promises. Our jointly deciding that I will φ gives you a particularly strong basis for forming an expectation that I will φ. First, it (normally) puts me under an obligation to φ, one that I owe *to you* in particular (§2.1). Our joint decision thus generates a new, and weighty, reason for me to φ. The more resolute our joint decision, the stronger this reason will be: which is why joint decisions that aim to provide assurance will typically be quite resolute (see §3). Second, the norms of joint practical deliberation require me to treat our decision that I will φ as I do my own intentions, excluding the possibility of not φing from my future deliberations and not being over-ready to reconsider this plan (§2.2). Finally, since joint decisions cannot be retracted unilaterally, I cannot simply change my mind about whether to φ without violating my obligation to you – even if I give you fair warning. If I want to retract our joint decision, I must propose to do so by means of propose-and-ratify, thereby giving you the final say over whether our joint decision remains in force (§2.4). These factors combined make joint decisions a powerful tool for giving others assurances about what we will do.

Thus the deliberative theory easily captures the important fact that promises provide assurance to their addressees. And it does so while keeping promises clearly distinct from testimony regarding one's future actions. Proposing to jointly decide that I will φ is clearly different from telling you or otherwise leading you to believe that I will φ. I can coherently tell you that I will φ while explicitly denying that I am proposing to jointly decide that I do so.

Moreover, the deliberative theory makes it easy to see why promises and testimony have such different normative upshots. Through the lens of the deliberative theory, we can see the distinction between promises and testimony regarding one's future actions as just the interpersonal analogue of the intrapersonal distinction between *intention* and *belief* regarding one's future actions. Consider the difference between my *intending* to finish this paper by May and my *believing* that I will do so. If I intend to finish the paper by May, I am committed to making this the case. While if I merely believe that I will do so, this does not imply any commitment to

85

action. Imagine that, midway through April, I notice that I am not on track to finish the paper before May. If I *intend* to finish the paper by May, I need to adjust my behavior to make this happen – say, by working longer hours. While if I merely believe that I will finish the paper by May, I need only adjust my *beliefs*. Seeing that I am not on track to finish by May, I may simply come to believe that I will finish by June, and not adjust my behavior at all.

The same contrast applies in the interpersonal case. If I promise you that I will finish the introduction to our coauthored paper by May, the result is a joint intention between us to the effect that I will do so. Like its individual analogue, this joint intention commits me to adjusting my behavior so as to make sure I finish the introduction by May. In contrast, if I merely *tell* you that I will finish the introduction by May – explicitly clarifying that this is a prediction, not a promise – then the result of my testimony is not a joint intention, but instead something like a shared belief. So if it begins to look like I won't finish the introduction by May, then all I need to do is adjust our shared beliefs to reflect this fact – say, by warning you that the introduction is taking longer than expected. The different obligations attending promises and testimony are thus a direct consequence of intention and belief's different directions of fit: we change the world to fit our intentions, while we change our beliefs to fit the world. Promises bind us to *act*, not just to warn, because they result in *decisions*, not just in beliefs.

4.3. The normative powers theory

As both conventionalism and the expectation theory have come under criticism, a third view has gained prominence: the *normative powers theory* (Raz 1972; 1977; Shiffrin 2008; 2011; Owens 2006; 2012). The normative powers theory rejects the idea that promissory obligation needs to be explained by appeal to a more general moral principle like the principles governing our participation in social practices or our care for others' expectations. Instead, the obligation to keep promises is "morally fundamental" and "derive[s] directly from the expression of an individual's will to be bound" (Shiffrin 2008: 482).

The normative powers theorist offers an alternative strategy for explaining promissory obligation. We are obligated to keep our promises, she argues, because it is *in our interests* to have the ability to undertake obligations by making promises. The ability to bind ourselves by promising is essential to human social life. If we lacked this capacity, we would be much worse off. On the normative powers view, the fact that having the ability to make binding promises is in our interests is sufficient to make it the case that we *do* have this ability.[26] Morality conspires to give us the tools that we need to live together, and one of these tools is the power to promise.

This explanatory strategy has traditionally been combined with a particular hypothesis about the interests that promises serve. In principle, the interest-based explanation just sketched could be adopted by someone who thought, with the expectation theorist, that the primary function of promises is to give others assurance. But most normative powers theorists are united in rejecting this idea (Raz 1972: 99; Owens 2006). The point of promising is not just to give others reliable information about what one will do: it is to give others *authority* over what one will do (Owens 2006). When I promise you that I will φ, I give you the authority to demand that I will φ and to rebuke me if I fail to do so. Moreover, I give you the authority to decide whether I am bound to φ by deciding whether to release me from my promise. These effects are often described in the language of *rights*: when I promise you that I will φ, I transfer my right to decide whether I will φ to you (Shiffrin 2011: 156; though see Owens 2014). The power to transfer authority in this way plays an important role in making valuable forms of human relationship possible (Shiffrin 2008). These benefits of transferring authority, in turn, ground the normative power to promise.

Understood this way, the normative powers theory is in fact compatible with the deliberative theory: the two could be adopted together. On this combined 'deliberative normative powers theory', the normative power to promise arises from the wider normative power to make binding joint decisions. Our obligation to abide by our joint decisions, in turn, is

---

[26] This argument strikes some as strange (myself included), as it seems to take the form "wouldn't it be nice if p, therefore p" (Enoch 2009). But many have argued that this strange form of inference is in fact valid within the moral domain (Kamm 1992; Nagel 1995; Enoch 2009; Preston-Roedder 2013; see van Someren Greve 2011 for criticism).

to be explained by appeal to our interest in being able to morally bind ourselves to the results of our joint deliberation. One (perhaps *the*) interest that the power to make binding joint decisions serves is to enable us to give one another authority over our actions. When we jointly decide that I will φ, I give you the authority to demand that I will φ and to rebuke me if I fail to do so. Moreover, I give you the authority to decide whether I continue to be bound to φ by deciding whether to agree to retract our joint decision. In fact, on the deliberative theory, there is a literal sense in which my promisee decides what I will do: the joint decision that determines my action is as much hers as it is mine. Thus the normative powers theorist can hold that promises transfer authority *by means of* making a joint decision.

Of course, the normative powers theorist could also deny the deliberative theory, claiming that promises are a *sui generis* normative power, not reducible to the power to make joint decisions. But in doing so, she would deny herself the explanatory resources that the deliberative theory offers. As we have seen, once we grant the bindingness of joint decisions, the deliberative theory gives us many of promises' features for free: their uptake requirements, validity conditions, and exclusionary effects on deliberation. These features can all be explained as consequences of the structure of joint practical deliberation. If she denies the deliberative theory, however, the normative powers theorist must instead write each of these features directly into her description of the normative power of promising. Then she must show why our interests most favor having *this exact kind* of normative power, with precisely the uptake conditions, validity conditions, and effects on deliberation that promising has. Perhaps this challenge can be met (see Owens 2012 for a formidable attempt). But if she instead adopts the deliberative theory, the normative powers theorist's task becomes much simpler. All she needs to do is show why it is in our interest to invest the results of our already-existing practice of joint practical deliberation with the force of moral obligation. Once we have that, the rest of promises' features will follow.

5. The redundancy argument

Here's what I've argued so far. If we are able to decide together what to do, then we must have speech acts by means of which we propose joint decisions. One of the speech acts we need is a proposal to make a joint decision about what the speaker will do by means of the propose-and-challenge method: what I have called an 'I-proposal'. Given plausible assumptions about the structure of joint practical deliberation, we can draw a number of conclusions about the properties such proposals must have. Together, these conclusions show that I-proposals have *exactly the same properties as promises*. When made and accepted in good faith, I-proposals have the same normative effects as promises: they give the speaker a strong *pro tanto* reason to perform the relevant action, typically obligating the speaker *to* the addressee to do so (§1.1, §1.2) and making it inappropriate for her to seriously consider doing otherwise (§1.3). For I-proposals to bind in this way, however, certain conditions must obtain: the proposal must be recognized and not challenged by the addressee (§1.4); the addressee must not have released the speaker from her obligation (§1.5); and the addressee must not have deceived or coerced the speaker into making the proposal (§2). These are the very same conditions that are required for promises to have their normative force. So an I-proposal has the same normative effects, under the same conditions, as a promise. The best explanation of these systematic similarities between promises and I-proposals is, I submit, that promises *are* I-proposals.

In offering this abductive argument, I have focused primarily on highlighting the deliberative theory's attractions rather than objecting to its opponents. However, we are now in a position to give an argument that challenges alternative theories more directly.

Say you accept that people engage in joint practical deliberation, and that one of the speech acts required for this activity has the very same properties that common sense ascribes to promises. If you accept this much, it is hard to go on and deny that promises *are* these proposals in joint practical deliberation. Doing so would commit you to the existence of two distinct speech acts with indistinguishable properties: I-proposals and promises. This view makes it mysterious why we have promises at all. I can obligate myself to you to φ by proposing to you that we decide

89

together that I will φ. So why would I need a second, distinct way of undertaking obligations to you with exactly the same normative upshots? Such a speech act would be redundant, unnecessary, even epiphenomenal. This is the status we must assign to promises if we deny the deliberative theory. But promises are *not* redundant in this way. Therefore, the deliberative theory must be true. Call this *the redundancy argument.*

We could, of course, avoid the argument's conclusion by simply accepting that promises are redundant. But this is deeply unattractive. Not only would this undermine the importance of promising — if promises are redundant, why should we care about them? — but, more importantly, it would saddle our theory with new and difficult mysteries. Why do promises exist, if their central functions can all be fulfilled by the more general practice of joint practical deliberation? If I-proposals are an additional speech act, separate from promises, then why don't we have a word for them in ordinary language? How has this entire category of speech act, at least as normatively significant as promises, escaped the notice of both ordinary language and philosophical inquiry? And why do these two separate speech acts, promises and I-proposals, have such eerily similar properties? Is it just a cosmic coincidence? The only way to avoid these awkward questions is to accept the deliberative theory.

Let's briefly consider how the redundancy argument applies to the three theories of promising we considered in section 4.

Begin with classical *conventionalism*, which says that promises are moves in the social practice of making and keeping promises (§4.1). Once we understand the nature of joint practical deliberation, we can see that all of the functions served by this promising-specific practice could be performed equally well by the wider practice of joint practical deliberation. If we already have the practice of joint practical deliberation, why would we need a separate practice of making and keeping promises? If we do have such a practice, it is redundant.

Now consider the *expectation theory*, on which to promise to φ is to offer your promisee a special kind of assurance that you will φ (§4.2). This special assurance is based in the promisee's knowledge that you are morally obligated, and specifically obligated *to him*, to φ. But if we engage

in joint deliberation, I can offer you exactly this kind of assurance by proposing that we jointly decide that I will φ. Why then would we need a distinct speech act of promising to provide a second way of offering the same assurance? If we do have such a speech act, it is redundant.

Finally, consider the *normative powers theory* (§4.3). On this view, to promise is to exercise a normative power to obligate oneself by declaration, granting one's promisee the authority to demand that one perform the promised action and to release one from this obligation if he wishes. But I can give you exactly these kinds of authority by simply proposing that we decide together that I will φ. Why, then, would we need a *sui generis* normative power of promising, if we can shape the normative landscape in exactly the same way by proposing joint decisions? If we do have such a normative power, it is redundant.

In sum, if one accepts that we engage in joint practical deliberation, and thus that we can propose to make joint decisions regarding our own actions by means of the propose-and-challenge method, it is strange to maintain that promises are a further, distinct speech act. Such a position is comparable to that of a nineteenth-century physicist who, acknowledging that mean kinetic energy exists and has all of the same observable features as temperature, still maintains that temperature is a further, distinct property. It is hard to see what could motivate such a view other than theoretical inertia. It is far more elegant and plausible to conclude that there are not two things here, but one: promises and I-proposals are one and the same.

We thus have a general argument against all alternatives to the deliberative theory: these theories all render promising redundant. The only way to escape this argument is to deny that there exist proposals in joint practical deliberation that have the same properties as promises. But that is the claim that the whole of this paper has been occupied with defending.[27] To deny this

---

[27] Even if you think promises are different from I-proposals in some respect, the redundancy argument should still give you pause. Suppose you are unconvinced by some part of sections 1 and 2: perhaps you think that insincere promises are valid while insincere I-proposals are not (§2.4), or that the power to release given by promises is more asymmetric than the power to release granted by I-proposals (§1.5). Still, if you grant that I-proposals share promises' most important features – they create obligations, give assurance, can be reversed with the addressee's consent, and are invalidated by coercion and deception – it seems to follow that I-proposals can serve all the major functions we attribute to promises. So, you face a choice: should you accept that promises are redundant, or instead question your commitment to thinking that promises are different from I-proposals in this particular respect? I think you should

claim, one must contest the theory of joint practical deliberation from which it follows. So, if the deliberative theory is false, that must be because the account of joint practical deliberation on which it rests contains some hidden flaw. For if we accept that joint practical deliberation exists and has roughly the shape I have described, the conclusion that promises are moves within this activity seems almost inevitable.

choose the latter: it seems easier to accept that common sense is misleading about one of promises' peripheral features than to swallow the conclusion that promises are redundant.

Chapter 3

Offers, Agreements, and Requests

In this chapter, I build a cumulative case for the deliberative theory by extending it to offers, agreements, and requests. I will not treat these three phenomena in the same detail as I did promises, partly due to space, partly because there is far less literature on them. Instead I will emphasize the deep similarities between these acts and promises. While the deliberative theory predicts these similarities straightforwardly, other theories are hard-pressed to explain them.

1. Offers

1.1. Similarities between offers and promises

We begin with offers:

The Deliberative Theory of Offers: For S to make an offer to A that she will

ϕ *just is* for S to propose to A, by means of the propose-and-ratify method, that

they make a joint decision to the effect that S will ϕ.

On the deliberative theory, what distinguishes offers from promises is their method of evaluation: while promises are meant to be evaluated by propose-and-challenge, offers are meant to be evaluated by propose-and-ratify. This captures the most salient difference between these speech acts: offers require explicit acceptance to come into force, while promises do not. This is why offers often take the form of questions: "Would you like for me to drive you home?" Offers call for a *response* in a way that promises do not. If you don't explicitly accept or reject my offer, it will hang in the air, unresolved – and eventually, I'll take your silence as a tacit rejection. Promises do not require a response to be complete in the same way. If you are silent after I make a promise, I'll normally take your silence as acceptance, and the promise will be in full force.

Beyond their different methods of evaluation, however, the deliberative theory holds that promises and offers are the same. When valid and accepted, both speech acts bring into force a joint decision that constrains only the speaker's actions. Since the normative effects of promises

derive from the normative effects of joint decisions, the deliberative theory predicts that offers, by bringing a joint decision into force, will have the same normative effects as promises. And since the validity conditions of promises derive from the conditions for good faith deliberation in general, the deliberative theory predicts that offers will share promises' validity conditions as well.

Common sense seems to agree with the deliberative theory on these points. Suppose I offer to pick you up at the airport and you respond, "Thanks, that'd be great." I am now bound to pick you up at the airport in the same way I would have been if I had promised to do so. Once you accept it, my offer gives me a strong *pro tanto* reason to pick you up at the airport. Unless this reason is overridden by unexpected circumstances, I will be obligated to pick you up, and specifically obligated *to you*. If I leave you stranded at the airport, you will have special standing to complain; if I look likely to renege on my offer, you can demand that I honor it. My offer also constrains my deliberation in the same way as a promise. Once you accept my offer, I should treat it as decided that I will pick you up at the airport, and not seriously consider any plans that conflict with this. And, as with promises, you (and you alone) have the power to release me from the obligation my offer created: "Actually, Ellen offered me a ride, so you don't have to pick me up after all."

Finally, just as coerced and deceived promises are not binding, so too with coerced and deceived offers. Consider these variations on our earlier examples:

> **BANK 2**: Bert holds a gun to Anna's head and says "OFFER TO GO TO YOUR BANK AND FETCH ME BACK A THOUSAND DOLLARS OR I'LL SHOOT YOU!" Anna says, "Yessir: I hereby offer to go to my bank and fetch you a thousand dollars." Bert says, "Thank you for your gracious offer; I accept."

> **VAN GOGH 2**: Cora makes an offer to Diego: "I offer to pay you $1000 for your Van Gogh." Diego says, "Wonderful! I accept." Cora sets off to the bank to get $1000 to pay Diego. On the way, she learns that the painting is forged, and Diego knew this was so.

Are Anna and Cora any more bound by these offers than they are by the promises in our original cases? I think not. Bert's (strangely polite) coercion renders Anna's offer invalid; Bert would not

94

be wronged if she did not give him the money. Diego's deception voids Cora's offer: he would not be wronged if she did not pay him for forgery.

It is indicative of how similar offers and promises are that philosophers rarely bother to distinguish them. But given that offers and promises are distinct speech acts, their similarities call out for explanation. Why do these two speech acts have such similar normative effects and validity conditions, differing only in their requirements for uptake? The deliberative theory answers this question straightforwardly: offers and promises are so similar because they are both proposals to make joint decisions about the speaker's actions; they are distinct because they propose this joint decision by means of different methods.

For those who do not accept the deliberative theory, life is not so easy. They will need to provide some alternative explanation of why offers and promises are so similar. The most obvious strategy is to claim that offers and promises are not distinct speech acts after all: offers are just a special type of promise. Let us give this view a closer look.

1.2. Is an offer a conditional promise?

The main feature that distinguishes offers from promises is that they need explicit acceptance from their addressees to come into force. If we want to claim that offers are a special kind of promise, we need to explain why they are different in this way.

The most plausible explanation seems to be: offers are *conditional* promises. When I offer to pick you up at the airport, what I am doing is promising to pick you up at the airport *if* you say you want me to. It is as if I had said, "I promise you this: if you say you want me to pick you up at the airport, I will do so." To generalize, the view we are considering is:

The Conditional Promise View: for S to make an offer to A that she will ϕ is

for S to promise A that, if A signals that she wants S to ϕ, then S will ϕ.[1]

---

[1] I use the unspecific term 'signal' to leave it open exactly *what* an addressee must do to count as accepting an offer.

The conditional promise view can easily explain why offers have the same normative effects and validity conditions as promises: because they *are* promises. But, I shall argue, the conditional promise view has trouble explaining the intuitive *differences* between offers and promises.

Here are two commonsense observations about promises. First, promises do not *expire*: there is no time limit after which a promise ceases to have its force. If, when we are young and sprightly, I promise you that I will come to your 80th birthday party, then my promise is made no less binding by the fact that it will be decades before I have to fulfill it. When your 80th comes around, I will have to make good on my word. Second, promises cannot be *retracted* unilaterally. I cannot simply take back my promises when I don't want to fulfill them anymore. I can ask you to release me from my promise, but I cannot take back my promise without securing your express permission to do so.[2]

Both of these commonsense points apply equally to conditional promises. Suppose you are playing the Pac-Man machine at our favorite bar. As the reigning champion, I challenge you to beat my high score. I say, "I promise you this: if you beat my high score on Pac-Man, I will buy you a beer." I thereby undertake an obligation to buy you a beer if you beat my high score. As with other promises, this obligation does not have an expiration date. If, many years from now, you beat my Pac-Man score, you can tell me to pay up. "You promised, remember?" I might be a bit chagrined that you are holding me to my promise so many years later, but I am nonetheless still bound by it. Second, I cannot revoke my promise unilaterally. If I see you racking up ever-higher scores in Pac-Man, I can't say, "Never mind, I take it back." To do so would not just be bad sportsmanship; it would be misunderstanding the very nature of a promise.

These points may seem so obvious as to not need saying. All the better. For neither of these points holds true for an offer that has yet to be accepted. On the conditional promise view, an offer that has yet to be accepted is simply a conditional promise whose condition has yet to be

---

[2] On the deliberative theory, this is due to an asymmetry in the methods by which promisors and promisees can propose to retract the joint decision brought into force by a promise: while the promisee can propose retraction using propose-and-challenge, the promisor can only do so by means of propose-and-ratify (see Ch. 2, §1.5).

fulfilled. If I make an offer and you say, "I'll think about it," I am in the same position as when I promise to buy you a beer if you beat my high score: I have undertaken a conditional obligation, but the condition has yet to obtain. Thus if offers are conditional promises, we should expect that unaccepted offers do not expire and cannot be unilaterally revoked. But this is not the case.

Suppose I offer to give you my antique bookcase. "I have too much stuff," I say. "Would you like to take it off my hands?" You aren't sure whether to accept my offer: the bookcase is nice, but you're not sure you have space for it. "I'll think about it," you say. Then we both forget about it. Several months pass, until one day, you discover the perfect spot to put the bookcase, and decide that you want it.

When you came to ask for my bookcase, what would you say? The natural thing to say would be something like: "Hey, is that offer still on the table?" or, "Do you still want to give away that bookcase?" You wouldn't simply say, months later, "I'll take the bookcase." That would be presumptuous. You should ask whether my offer is still on the table because there's a real chance that it *isn't*. I might have given the bookcase to someone else by now, or decided that I want to keep it after all. And importantly, if either of these things is true, I have not done you wrong: I was not obligated to keep the bookcase available indefinitely, in case you decided you want it. This shows that offers *expire*: if an offer is not accepted, then after a certain (vague and context-dependent) period of time, the offer is no longer available to be accepted. If you want my bookcase, you'll have to ask if I am willing to offer it again.

This is in stark contrast with conditional promises. Say that, months after I promise to buy you a beer if you beat my high score, you finally do it. Triumphant, you wouldn't then say, "Hey, is that promise still on the table?" You would say: "I'll take that beer now." That's because conditional promises, unlike unaccepted offers, do not expire.

For a closer comparison, suppose that instead of offering you the bookcase, I said, "I promise you this: if you ask for the bookcase, I'll give it to you." Then the situation would be quite different when, months later, you decide you want it. You could ask for the bookcase directly, knowing that I am still bound to give it to you. And if I had given the bookcase to

someone else in the meantime, then I would have done you wrong. My conditional promise committed me to keeping the bookcase available in case you asked for it.

Let us turn to the second point of contrast between offers and conditional promises: before they are accepted, offers can be *retracted* unilaterally. Say, again, I offer to give you my bookcase and you say, "I'll think about it." But then someone else offers to pay me a large sum for the bookcase. I can call you up and say, "Sorry, something came up – I can't give you my bookcase after all." I am perfectly within my rights to do this. Before you accept my offer, I am free to revoke it at any time.[3]

This is not the case for conditional promises. As mentioned above, I cannot simply take back my promise to buy you a beer if you beat my Pac-Man score, even before you beat it. And the same holds if I promise to give you the bookcase if you ask for it. This promise commits me to holding the bookcase for you; if I want to give it to someone else, then I have to ask you to release me from my promise first.

That offers have these two features – expiration and revocability – appears to be standard doctrine in contract law. After a contract is offered, the offeree must accept it within a reasonable (sometimes explicitly stated) period of time, or else it lapses, becoming void. And before the offeree accepts a contract, the offeror may legally revoke it at any time. The following passages are illustrative of this consensus:

> An offer having been made will continue until it is ended by a lapse or a revocation. The offerer may, by informing the offeree of a change of mind, terminate the offer by thus destroying the expectation which his offer has aroused and for which he is responsible (Oliphant 1920: 204-205).

---

[3] Here it is important to distinguish the question of whether I have the *authority* or *power* to revoke my offer from whether I *ought* to do so. There might be strong reasons why I should not revoke an unaccepted offer: for instance, you might be relying on the expectation that you can have my bookcase if you want. For this reason, perhaps I should give you another chance to accept the offer before I take it back and give the bookcase to someone else. The important point is that offers *can* be revoked unilaterally, even if they should not be. Before you accept my offer, I have the power to take it back without asking for your permission to do so. When I say, "Sorry, I can't give you the bookcase after all," I thereby successfully revoke my offer. This action is perfectly intelligible, even if it is rude. But it is *impossible* to revoke a promise in the same way – I simply don't have the power to do so. The closest thing I can do is inform you that I will break my promise, which is a very different thing from revoking an unaccepted offer.

It is elementary that an offer not under seal and without consideration may be revoked at any moment prior to its acceptance (McGovney 1914: 644).

In order to lock in the deal, B [the offeree] must accept the offer in a proper manner, and within a proper period of time – that is, before S [the offeror] has revoked the offer, and before the offer has terminated for some other reason. The standard rule holds that, even if S never revoked her offer, B's power of acceptance ends if he does not accept the offer within a 'reasonable time' (Craswell 1996: 512).

An offer having been made and a power having been thereby created, how long will this power continue to exist? The offeror is the creator of the power, and before it leaves his hands he may fashion it to his will. Such is the present decree of society. If he names a specific period for its existence, the offeree can accept only during this period. If the offeror names no period whatever, the power will be held to exist for a reasonable time, to be determined as a fact by the court and dependent upon the circumstances (Corbin 1917: 183).

Can the conditional promise view be modified to accommodate these facts? Yes, but at a cost. To accommodate expiration, we could say that offers are really conditional promises of the form, "if you signal that you want me to ϕ *within a certain period of time*, I will ϕ." To capture revocability, we would have to add an opt-out clause to the conditional promise. On the resulting view, when I offer to give you my bookcase, I am making the following complex promise: "I promise you this: I will give you my bookcase *if* you ask me for it within a reasonable period of time, *and if* I do not change my mind at some point before you ask me for it."

This is a strange sort of promise. Imagine if I made a similar promise in the Pac-Man case: "I promise you this: if you beat my high score in Pac-Man within a reasonable period of time, I will buy you a beer, *unless* I change my mind at some point before you do." That's not a promise – that's a joke. To say "I promise, unless I change my mind" is to undercut your promise so dramatically that it is barely intelligible as a promise at all.

Here's a less flat-footed objection. The conditional promises view can indeed capture offers' intuitive features by squeezing more implicit clauses into the relevant promises' content. But in doing so, it makes it mysterious why we make offers at all, not to mention singling them out in ordinary language as their own special category. The subset of promises that the view identifies with offers – conditional promises with a time limit and an opt-out clause, both of which only apply before the condition is met – is so specific and gerrymandered that it is surprising that we make promises of this kind at all. Why, for instance, don't we add opt-out clauses to our unconditional promises, or allow our conditional promises to lapse after they are accepted but before they are performed? Why is this one specific kind of conditional promise, with its complex set of implicit conditions, performed so commonly that it has its own entry in the dictionary? Perhaps it would be worth swallowing these mysteries if the conditional promise view was the only account of offers available. But it isn't.

On the deliberative theory, offers are not conditional promises. A conditional promise, when valid and accepted, brings into force a binding joint decision with conditional content. When I promise to buy you a beer if you beat my Pac-Man score, the result is a joint decision between us to the effect that if you beat my Pac-Man score, I will buy you a beer. Just like any other joint decision, this decision cannot be revoked unilaterally and does not expire.

If I make an offer, on the other hand, then until you accept it, no joint decision is made, conditional or otherwise. All I have done is *propose* a joint decision. Since proposing is an act I perform individually, rather than a commitment we jointly make, I can retract my proposal at any time, simply by indicating that I no longer stand by it. In this way, offers are more like assertions than promises: if I change my mind about an offer, I can take it back by simply saying, "Never mind."

And there is no reason to think that, once I propose a joint decision, my proposal will remain on the table indefinitely, awaiting your response. Deliberative proposals remain open only so long as it is common ground between us that they are open. If I make a proposal and you say, "I'll think about it," then after a certain amount of time the proposal will no longer be an

100

active question in our joint deliberation, simply because we are no longer treating it as such. If you want to take it up again, you will have to propose the joint decision yourself, or see if I am willing to make the proposal again. In this way, offers are like factual questions: if I ask you what your summer plans are, there's only a certain period of time in which your saying "I'm traveling to Spain" will count as an answer to my question.

The conditional promises view can accommodate the observation that unaccepted offers can expire and be revoked, but only by building these features in after the fact — and at a steep theoretical cost. In contrast, the deliberative theory independently predicts that offers will have these exact features. It seems clear to me which account is preferable.

I conclude that offers are not conditional promises. Offers and promises are distinct speech acts, neither reducible to the other. The deliberative theory is well placed to explain both the similarities between these speech acts and their differences — including the fact that, before they are accepted, offers can expire and be revoked. And as the strategy of reducing offers to promises appears unworkable, the opponent of the deliberative theory is back at square one. She is left with an unsolved mystery: if offers are not promises, then why are they so similar to them?

1.3. Differences between offers and promises

We have already covered the central differences between offers and promises: first, that offers require explicit acceptance to come into force, while promises do not (§1.1); and second, that unlike promises, offers expire and can be revoked before they are explicitly accepted (§1.2). Here I want to explore one further set of differences: the different conditions under which promises and offers are socially appropriate.

For example, it would be inappropriate for you to promise, unbidden, to read an early draft of a paper I am working on, as this seems to presume that I want you to read it. But it would be perfectly appropriate for you to *offer* to read my draft, since that gives me the option of turning you down. Conversely, it would be strange, even offensive, for me to *offer* to be honest to my partner. Phrasing it as an offer seems to imply that I'm doing her a favor. But it is perfectly

appropriate for me to promise my partner that I will be honest with her (even if this is only reaffirming a commitment I've already undertaken).[4]

We can explain these differences in appropriateness by looking at when it is appropriate to employ the propose-and-challenge vs. propose-and-ratify methods. When I introduced these two deliberative methods (Ch. 1, §3.1), I pointed out two main differences between them. First, the propose-and-challenge and propose-and-ratify methods guard against different kinds of *error*. By making it harder for a joint decision to come into force, the propose-and-ratify method guards against *false positive* errors: the error of making a joint decision that is not warranted. The propose-and-challenge method, in contrast, makes it relatively easy for a proposal to come into force. Thus it guards against *false negative* errors: the error of failing to make a joint decision that is warranted. Second, our two methods create different balances of power. Propose-and-challenge puts greater deliberative power in the hands of the speaker, while propose-and-ratify gives the deliberative power to the addressee.

Taking these factors into account, we can predict the conditions under which promises will be more appropriate than offers, and vice versa. Here I review four such predictions.

(i) *Promises are appropriate only when the speaker justifiably takes herself to know that the promise is warranted. Conversely, only offers are appropriate when the speaker does not know whether the promise is warranted.* If the speaker is uncertain about whether the decision she is proposing is warranted, then there is significant risk of a false positive error. So, she ought to use the deliberative method that guards against false positive errors: the propose-and-ratify method. When you are uncertain as to whether your proposal is warranted, then, you should phrase it as an offer. Conversely, you

[4] It is hard to nail down the evaluative import of these judgments of appropriateness and inappropriateness. To say that it is inappropriate to make a promise or offer in a certain scenario is not necessarily to say that the promise or offer is unwarranted, i.e. that it fails to conform to the norm of joint decisions. My offer to be honest to my partner is certainly *warranted* – our shared reasons decisively support the joint decision it proposes – but it is still inappropriate. And your offer to read my draft is perfectly appropriate, even if the joint decision it proposes turns out to be unwarranted (since I don't want your feedback at this stage of the paper). Judgments of appropriateness and inappropriateness won't coincide with judgments of agents' all-things-considered reason to make a speech act either. I might have overriding reason to make an inappropriate promise or offer – say, if someone holding a gun tells me to do so. I don't have a theory of what exactly it means to be appropriate or inappropriate, but I think we have a firm enough grip on these notions to leave them unanalyzed.

ought to make a promise only when you are *not* in doubt as to whether the proposal you are making is warranted.

This prediction fits our judgments. It is inappropriate for me to promise, unbidden, to read your draft because I do not know whether you want me to do so, and so do not know whether we would be warranted in deciding that I will read your draft. I should offer to read your draft instead, ensuring that the decision won't be made unless you confirm that it is warranted. Generally, an important function of offers is to give someone the option of taking our help when we aren't sure whether they want it. This allows the addressee to determine whether the proposed decision is warranted and accept or reject it on that basis. To *promise* help when you aren't sure it is wanted is off-base precisely because it doesn't allow for this epistemic check.

However, when you take yourself to already know that your addressee wants you to do something (perhaps because they've told you, perhaps because it's obvious), then promising to do it can be perfectly appropriate. Intuitively appropriate promises – promising to be honest to my partner, to pay you back, to be home in time for dinner – seem to fit this mold.

(ii) *Ceteris paribus, offers are more appropriate than promises, since they give the addressee more deliberative power.* Even when you are in a position to know that a joint decision is warranted, that doesn't always mean that propose-and-challenge is appropriate. Given the asymmetrical power involved in our deliberative methods, the choice of whether to promise (giving yourself more power) or offer (giving your addressee more power) has significance that goes beyond the epistemic considerations just cited.

Ceteris paribus, it seems better to give the deliberative power to your addressee. Now, *ceteris* are not always *paribus*: there are situations in which you have good reason to take power into your own hands (see (iii) and (iv) below). But the default choice should be to offer rather than promise (and, similarly, to request rather than command; see §3.3 below). Why? Because there is a natural temptation to take power for oneself, both in order to skew the results of deliberation in one's favor and because one tends to trust one's own judgment more than the judgment of others.

103

These self-serving biases are the perennial foe of equal cooperation. So, when there is little cost to doing so, it is better to cede deliberative power to others.

This seems to me to fit our everyday practice. By my observation, at least, it seems that offers are much more common than promises. People often make an offer even when they are, epistemically speaking, in a position to promise. An example: my colleague and I almost always read one another's work. It's common knowledge between us that if one of us has written a draft, he would like the other to read it. Still, when he tells me he's finished up a draft of a new paper, I don't just say straight-out, "I'll read it, I promise." I ask, "Would you like me to read it?" I know the joint decision that I will read my colleague's draft is warranted, but it is still more polite to offer rather than promise to read it. Why? Because by doing so, I give my colleague the deliberative power rather than taking it for myself.

(iii) *When the addressee needs assurance that the speaker is motivated to perform the proposed action, promises are more appropriate than offers.* One case where it is appropriate to take the deliberative power for oneself is when one needs to demonstrate one's commitment to the proposed decision. If I have the power over whether a joint decision is made, then proposing it counts for more. By using propose-and-challenge, I ensure that my proposal will come into force by default, rather than just if my addressee accepts it. If my addressee is less concerned about having deliberative power than she is about knowing that I am committed to the decision, then she might wish for me to affirm the decision more forcefully by using the propose-and-challenge method.

This characterizes many of the situations in which promises are made. Often, a speaker will promise to do something that she already intended to do, simply to reassure her addressee that she will do it. I say "I'll be home in time for dinner, I promise" because my addressee is worried that I won't make it home in time, and needs to be assured that I will take the necessary measures to do so. I promise my partner that I'll be honest in order to express my commitment to doing so, and thus provide her with reassurance.

When the addressee needs to be assured of the speaker's commitment, offers seem inappropriate. If your willingness to do something is in doubt, *offering* to do it seems to show that

104

you're holding out hope that the addressee will turn your offer down. Even if you know that the offer will be accepted, phrasing it as an offer seems to say, "I'll do this, *if you insist*." By giving the addressee the last word over whether you will perform the action, you make it the *addressee's* responsibility to decide whether you do so, rather than taking that responsibility for yourself. But what the addressee wants is for *you* to decide to perform the relevant action, and thereby demonstrate your commitment to doing so. The best way to do this is to take control into your own hands by *promising* to do it.

This difference has implications for the relative force of promises and offers. If promises are more likely than offers to be made in order to give the addressee assurance, then we can predict that promises will tend to result in more resolute joint decisions than offers do (see Ch. 2, §3). For, the more resolute a joint decision, the stronger assurance it will provide. The upshot is that promises will tend to generate stronger obligations than offers do. This seems right: making a promise seems to be a more serious undertaking than making an offer. Rather than furnishing an objection to the deliberative theory, however, this intuitive contrast is predicted by it: it is a direct result of the different functions of the propose-and-challenge and propose-and-ratify methods.

(iv) *When the addressee is likely to reject a warranted proposal, promises are more appropriate than offers.* Many of us know people who turn down help even when they need it – out of pride, self-denial, or anxiety about burdening others. Those of us who have such people in our life quickly learn that *offering* them help is the wrong way to go. To use an example from earlier, suppose you have a proud but frail grandmother who needs help moving into her new home. If you say, "Would you like me to help you move?" she will inevitably turn down your offer, and that will be the end of the conversation. If you reiterate the offer, your grandmother will continue to reject it. And once she has turned you down, showing up to help anyway would be disrespectful.

Much better to *promise* help: "Grandma, I'm helping you move." Your grandmother may still resist: "No, I'll be fine." But now the dynamic is different. All you need to do is counter her objections: "Nonsense. There's too much stuff for you to move alone!" Now, the onus is on your grandmother to provide a defensible challenge to your promise. If your promise is warranted, she

will have a tough time doing so, and eventually she will have to give in and accept your help. That is a good thing: she needs your help, after all!

Though, other things equal, it is better to err on the side of granting deliberative power to your addressee, this is one of the cases where other things are not equal. Your grandmother cannot be trusted to accept a warranted joint decision if it involves you helping her in some way. This doesn't necessarily mean that she is deliberating in bad faith; she may just be biased and stubborn. But the result of your grandmother's biases is that you are at high risk of the false negative error of failing to jointly decide that you will help her when you have decisive shared reason to do so. So, you should take the deliberative power out of Grandma's hands, using propose-and-challenge to push through the warranted decision that you will give her help. This is why you should *promise* help rather than merely offering it.

I am sure that the subject of when it is appropriate to promise versus offer is not exhausted by the observations I have made here. But I hope to have shown that, when we make this choice, we are guided by our tacit knowledge of the costs and benefits of the propose-and-challenge and propose-and-ratify methods.

2. Agreements

Now we turn to agreements:

The Deliberative Theory of Agreements: For S and A to make an agreement that they will φ *just is* for S and A to make a joint decision to the effect that they will φ.

This thesis sounds almost tautological: *of course* agreements are joint decisions! True enough. The interesting claim of this dissertation is not that agreements are joint decisions, but that they are not the *only* kind of joint decision. The deliberative theory's surprising thesis is that so many speech acts that do not appear to involve joint decisions – promises, offers, requests, demands, commands – are moves within joint deliberation as well. My main purpose in discussing

106

agreements is thus not to defend the deliberative theory of agreements itself, but instead to bolster the case for the other, more controversial parts of the deliberative theory.

I just argued that the opponent of the deliberative theory has trouble explaining the similarities between offers and promises, because offers cannot be plausibly reduced to promises. I will now make the same argument regarding agreements. While the deliberative theory explains the similarities between promises and agreements easily, these similarities pose an awkward question for other views. For, like offers, agreements cannot be reduced to promises.

It is uncontroversial that agreements share promises' central features (cf. Sheinman 2011). Suppose we agree to meet at Clover for lunch. Then this agreement puts both of us under the same kind of obligation we would have if we had each promised individually to go to Clover. We are each obligated to show up to Clover (barring some unexpected reason that overrides our commitment), and obligated to one another to do so. This agreement constrains both of our subsequent deliberations: it would be inappropriate for either of us to treat the question of whether to go to Clover for lunch as open for further deliberation. Finally, we can release each other from the obligations created by our agreement by retracting it: "Actually, could we take a rain check until next week?" "Sure, that's fine." Agreements thus share the pattern of normative effects that we've already seen in promises and offers.

Agreements, too, are invalidated by coercion or deception. Consider

BANK 3: Bert holds a gun to Anna's head and says, "LET'S MAKE AN AGREEMENT: YOU WILL GO TO YOUR BANK AND FETCH ME ONE THOUSAND DOLLARS, AND IN EXCHANGE I WON'T SHOOT YOU." Anna says 'Yessir.'

VAN GOGH 3: Diego proposes an agreement to Cora: "How about you go to your bank and fetch me one thousand dollars, and in exchange, I'll give you my Van Gogh." On the way to the bank, Cora learns that the painting is forged, and Diego knew this was so.

Bert's coercion and Diego's deception rob the resulting agreements of their normative force – just as they rendered the promises in BANK and VAN GOGH invalid.

The primary difference between agreements and promises is that promises bind only one person's actions, while agreements bind multiple parties' actions simultaneously. As we saw in Chapter 2 (§3), this leads to some subtle downstream differences between the two. First, since agreements bind the participants' actions symmetrically, their retraction conditions are also symmetrical: both parties will typically be able to propose to retract the agreement by means of the same method. Whether that method is propose-and-challenge or propose-and-ratify will depend on the situation – specifically, on whether the participants ought to make the agreement easier or more difficult to retract. This contrasts with promises (and offers), which, because they bind the participants' actions asymmetrically, have correspondingly asymmetrical retraction conditions. While the promisee can propose to retract the promise using propose-and-challenge, the promisor must use propose-and-ratify.

The second main difference between promises and agreements concerns the purposes for which they are typically made. As I argued earlier (Ch. 2, §3), promises almost always serve the function of strengthening the promisor's commitment to a course of action; for this reason, the joint decisions they propose will usually be very resolute. But while agreements *can* serve this committing function, they more often simply serve the planning function of coordinating both parties' plans to make joint action possible. Since this planning function does not require a high level of resolve, we can predict that the joint decisions resulting from agreements will have a wider range of resolve, being more or less resolute as the situation requires. This difference in resolve, I argued, explains the intuitive observation that promises typically generate stronger obligations than agreements.

But the similarities between agreements and promises are far more significant than these differences. Intuitively, agreements have the same kind of normative force, under the same conditions, as promises. This is just as the deliberative theory predicts. But for the opponent of the deliberative theory, who does *not* hold that promises result in joint decisions, the similarities between promises and agreements should be surprising. If promises do not involve joint decisions, but agreements are joint decisions, then why should the two be so similar?

As with offers, it seems that the best strategy available to the opponent of the deliberative theory is to try to show that agreements are reducible to promises in some way. The canonical way of doing this is to claim that agreements are *promise exchanges*: when we agree to meet at Clover, what really happens is that I promise you that I'll be at Clover, and in exchange, you promise me that you'll be there as well. But, as we shall now see, this reductive strategy cannot be made to work.

The hard work has been done for us already. In her classic paper "Is an Agreement an Exchange of Promises?" Margaret Gilbert argues at length that the answer to her title question is *no* (Gilbert 1993b). I find her arguments convincing, so I will review them here.

Gilbert begins by observing three features of agreements that any theory must capture:

(1) **Obligation:** "the agreement directly generates the relevant performance obligation for each of the parties" (630).

(2) **Simultaneity:** "In many cases, a given party will have no desire to take on some particular obligation before the other party has taken on a corresponding obligation ... [Agreements ensure] that obligations are undertaken simultaneously by the parties, even though one may speak or relevantly gesture before the other. No one is *obligated first*" (630).

(3) **Interdependence:** "the obligations of ... agreements are interdependent at least in the following way: if one party defaults on his performance obligation, the other ceases to have his original performance obligation" (630).

Gilbert then argues that no exchange of promises can have all three of these features. I do not have space to survey all of the proposals Gilbert considers, but a couple examples will illustrate the problems faced by the promise-exchange view.

Gilbert considers the following sample agreement: "Rita: 'Someone must walk Fido this afternoon, and Tibbles has to be groomed.' Peter: 'Why don't you walk Fido, and I'll groom Tibbles?' Rita: 'Fine!'" (627). Can we reconstruct this agreement as an exchange of promises?

First, consider the simplest analysis: Peter promises to groom Tibbles, and Rita promises to walk Fido (635).[5] This exchange fails the interdependence criterion, simply because pairs of promises are not normally interdependent. Say you promise me that you will read my draft by Friday, and in an unrelated event, I promise to look after your cat next week. You fail to read my draft in time. Am I off the hook for cat-sitting? It doesn't seem so. The fact that you broke your promise doesn't nullify mine. The same holds for Peter and Rita's promises: if one of them breaks their promise, this does not nullify the other. This simple promise exchange fails the simultaneity criterion as well: whoever makes their promise first is obligated before the other is.

We can build in simultaneity as follows. Suppose that Peter says: "I promise to groom Tibbles, if you promise to walk Fido." Then Rita says, "I promise to walk Fido" (642). Then Peter and Rita's obligations come into force at the very same time: once Rita makes her promise. Though this analysis yields simultaneity, it gets us no closer to interdependence. The result of Peter and Rita's exchange is still two independent promises: Peter has promised Rita that he will groom Tibbles, and Rita has promised Peter that she will walk Fido. Neither of these promises loses its binding force if the other promise is broken.

With a little fancy footwork, we can build in interdependence as well. Suppose that Peter says this: "'On condition that you make the very same ... promise, I promise to do my part in the following pair of actions if you do your part: Peter grooms Tibbles and Rita walks Fido.' Rita makes the same promise as Peter" (640). The obligations that result from this exchange are interdependent: if Rita does not walk Fido, then Peter is not obligated to groom Tibbles, and vice versa. And they also seem to be undertaken simultaneously: since Peter's promise is conditional on Rita's making the same promise, both Peter and Rita's obligations come into effect at the moment when Rita makes her promise.

However, as Gilbert points out, this exchange of conditional promises fails the obligation criterion: it does not obligate Peter to groom Tibbles or Rita to walk Fido. The obligations that

---

[5] The objections I present in the next three paragraphs are all Gilbert's; I am just paraphrasing them.

result from this exchange are conditional: Peter is obligated to groom Tibbles *if* Rita walks Fido, and Rita is obligated to walk Fido *if* Peter grooms Tibbles. One way for Peter and Rita to satisfy these obligations would be for Peter to groom Tibbles and Rita to walk Fido. But another, equally legitimate way for Peter and Rita to satisfy their obligations is for both of them to *do nothing*. So long as Peter does not groom Tibbles, Rita is under no obligation to walk Fido, and vice versa. Thus this exchange of promises fails to give Peter and Rita reason to do anything at all: they can both fulfill their obligations by sitting on the couch and neglecting their poor pups.

This final problem seems fatal to me. For it is hard to see how to make two promises interdependent except by making each promise conditional on the other's fulfillment. But mutually conditional promises can always be fulfilled by both parties doing nothing. Agreements are different. Though the obligations they generate are interdependent, agreements are not fulfilled when both parties break them. We both have reason to ensure that our agreement is honored: if we both break our agreement, then we have both failed to do what we had reason to do. Unlike mutually conditional promises, agreements do not treat joint inaction as normatively equivalent to joint fulfillment. This suggests that it is impossible in principle for two promises to imitate the normative effect of an agreement.

Let me briefly indicate how the deliberative theory captures the three features of agreements that promise-exchange theories cannot.

The obligation criterion is easy. Agreements generate obligations in the same way that promises do: by bringing a joint decision into force. If we retain the assumption that people are typically obligated to abide by their standing joint decisions (see Ch. 2, §1.1), then people will typically be obligated to abide by their agreements.

Both the simultaneity and interdependence of agreement-based obligations are explained by the fact that agreements involve a *single* joint decision concerning the actions of multiple agents. An agreement that we will ϕ brings into force a single decision *that we will* ϕ, not two separate decisions *that I will do my part in our ϕing* and *that you will do your part in our ϕing*. Thus the

various obligations that result from an agreement are all created at the same moment: the moment at which our joint decision comes into force.

To see why agreement-based obligations are interdependent, note that on the deliberative theory, our reason to keep our agreements derives from our reason to abide by our standing joint decisions. Recall from Chapter 1 (§1.1 and §1.2) that the content of a joint decision is a set of possible combinations of our individual actions. So, the content of our joint decision *that we will* φ is the set of possible combinations of our individual actions in which we both do our part in our φing. To conform to this joint decision, we must collectively keep our behavior within the bounds of this set. So, if one of us acts in a way that falls outside of this set – in other words, if one of us breaks our agreement – then that alone is sufficient to make it the case that we have not acted in accordance with our joint decision. Once you have broken our agreement, there is nothing I can do to bring our actions within the set of action combinations that we jointly decided upon. So, my reason to uphold our standing joint decisions no longer gives me any reason to do my part in our φing, since my doing so will no longer make it the case that our joint decision has been upheld. Hence interdependence. Crucially, however, so long as it *is* still possible for us to fulfill our joint decision, we both have reason to do so. Thus the deliberative theory avoids the conclusion that we can always fulfill an agreement by doing nothing.

This explanation of why agreements yield interdependent obligations does not extend to exchanges of promises. Say Rita promises Peter that she'll walk Fido, and Peter promises Rita that he'll groom Tibbles. The deliberative theory claims that these promises result in two separate joint decisions: a decision to the effect that Rita will walk Fido, and a decision to the effect that Peter will groom Tibbles. Since these are two separate decisions, the violation of one is compatible with the fulfillment of the other. If Rita violates their joint decision that she will walk Fido, that is unfortunate, but it does not affect Peter's ability to execute their joint decision that he will groom Tibbles. Thus the deliberative theory predicts correctly that while agreements create interdependent obligations, exchanges of promises do not.

112

Drawing on Margaret Gilbert's work, I have argued here that the opponent of the deliberative theory cannot explain the similarities between promises and agreements by reducing agreements to promise exchanges. It is more plausible to go the other way around and say, with the deliberative theory, that promises are a special kind of agreement: a joint decision concerning the speaker's actions. Those who refuse to take this route are left with another mystery: if agreements are not exchanges of promises, then why are they so similar to them?

3. Requests

3.1. Similarities between requests and promises

Our third and final topic in this chapter is requests:

The Deliberative Theory of Requests: For S to request of A that she φ *just is* for S to propose to A, by means of the propose-and-ratify method, that they make a joint decision to the effect that A will φ.

As with offers and agreements, we will begin by surveying requests' similarities to promises.

Suppose I ask you to read my draft, and you say, "Sure!" By accepting my request, you become obligated to read my draft. The obligation you undertake by accepting my request is much like the obligation you would have undertaken if you had promised to read my draft instead. The list should by now be familiar. By accepting my request, you gain a new strong *pro tanto* reason to read my draft, a reason that can be overridden but typically generates an obligation for you to read it. This obligation is directed: you owe it *to me* in particular to read my draft. Like promissory obligation, this obligation imposes constraints on your deliberation: it would now be inappropriate for you to treat the question of whether to read my draft as a matter for further deliberation. And as with promises, I alone have the power to release you from the obligation you took on in accepting my request: "Actually, I'll have a better draft soon, so don't worry about reading that one."

Like promises, offers, and agreements, requests are invalidated by coercion and deception as well. Consider a fourth (and final) variation on our favorite examples:

113

BANK 4: Bert holds a gun to Anna's head and says, "WOULD YOU PLEASE GO TO YOUR BANK AND FETCH ME ONE THOUSAND DOLLARS? IF YOU DON'T, I'LL HAVE TO SHOOT YOU." Anna says 'Yessir.'

VAN GOGH 4: Seeing Cora eye his Van Gogh, Diego says, "Would you take it off my hands? It's only $1000." Cora says, "Sure!" and heads to her bank to get the money. On the way, Cora learns that the painting is forged, and Diego knew this was so.

As in our other cases, Bert's coercion and Diego's deception rob the interaction of any binding force. By accepting Bert's request, Anna does not become obligated to go fetch him the money; by accepting Diego's request, Cora does not become obligated to pay for his forged Van Gogh.

Thus requests seem to create obligations with the same normative upshots as promises under the same conditions. The two speech acts seem to differ only in the subjects of the obligations they create: requests bind their addressees, while promises bind the speaker.

The deliberative theory has a ready explanation for these similarities. Requests are similar to promises because they have the same result: bringing a joint decision into force. When you ask me to read your draft, and I accept your request, the result is that we have jointly decided that I will read your draft. When I promise to read your draft, and you do not object, the result is the same: we have jointly decided that I will read your draft. This joint decision gives rise to an obligation with all the features listed above.

Requests are invalidated by coercion and deception for the same reason that promises are: coercion and deception are incompatible with good faith joint deliberation. While coerced and deceived promises are invalidated by the bad faith of the addressee, coercive and deceptive requests are invalidated by the bad faith of the speaker. But the mechanism is the same: if you are coercing or deceiving me, then you cannot be deliberating with me in good faith, and so our interaction cannot give rise to a binding joint decision.

Thus the deliberative theory predicts that requests and promises will have similar normative effects and validity conditions. The question, again, is how other views can explain

these similarities. Again, the reductive strategy – explaining requests in terms of promises – seems like the only viable route. So let us see if it can be made to work this time.

3.2. Are requests solicitations of promises?

The observation we are trying to explain is this: by accepting a request, one undertakes an obligation that is very similar to the obligation one undertakes by making a promise. Thus a simple explanation presents itself: accepting a request just *is* making a promise.

What, then, is it to *make* a request? It must be something like a promise solicitation: when I ask you to read my draft, I am trying to get you to promise to read my draft. How? We can't say that I am *asking* you to make a promise – that would be circular. The only non-circular interpretation of the idea that requests are promise solicitations I can think of is this: to make a request is to *express your desire* that your addressee promise to do a certain action.

This view does not face the same kind of direct counterexamples that defeated our other reductive proposals. After all, the deliberative theory agrees with this reductive view that the normative upshots of accepting a request will be indistinguishable from those of making a promise. While the two views agree on the effects of *accepting* a request, however, I want to suggest that the deliberative theory better explains the import of *making* a request.

The act of making a request changes the normative situation even before it is accepted. To see this, suppose that you and I are beginning a fledgling friendship: we are not mere acquaintances, not quite friends, but somewhere in between. I call you up and make a request: "Hey, I'm moving to a new apartment on Saturday, and I can't move my furniture alone. Any chance you could come over and help me out?"

My request puts you in a bit of a bind. On the one hand, giving up your Saturday to lift my heavy furniture would be beyond the call of duty even if we were closer friends. Given the liminal status of our friendship, it's a lot to ask. If I had casually mentioned that I was moving on Saturday, you would have felt no obligation to offer your help. And importantly, my request alone does not obligate you to help: you are still free to turn me down. On the other hand, my

115

request does seem to oblige you to give *some* reason for turning it down. If you had some prior plans, you could cite those. But if not – if, say, you were just planning to spend Saturday relaxing – you'll be in an awkward position. "I don't want to" is generally not a socially acceptable reason for rejecting a request.[6] You'll either have to accept my unreasonable request, or pretend you have some better excuse for turning it down.

Perhaps it is because we are aware of this fact that we are careful about what requests we make. Many of us have had the experience of agonizing over whether we can reasonably make a certain request – or, when we do make a request, trying to assure our addressee that she can turn us down. We don't want to put other people in the bind of having to respond to an unreasonable request, and so we often only make requests when we already know the answer will be 'yes.'

On the reductive view we are considering, these everyday experiences should be surprising. All I do when I make a request, on this view, is express a desire that you do a certain thing. But we normally do not feel obliged to provide a justification for failing to satisfy others' desires. If you say, "I want a chocolate cake," I will feel no pressure to offer a reason why I won't make one for you. Granted, we do often make requests *by* expressing our desires. When I say, "I want for you to read my draft," I both express a desire that you read my draft *and* ask you to do so. So the trouble for the reductive view isn't that we can't make requests by expressing desires. My complaint is more subtle: I can't see why requests would make their addressees feel obliged to offer a justification for turning them down if they are *nothing more* than expressions of desire. We normally don't owe others a justification for not satisfying the desires they express; why, on the reductive view, should requests be any different?

The deliberative theory better captures the phenomenology of requests. Using the resources of the theory, we can explain our example as follows. When I request that you help me move, I am not merely expressing a desire that you do so. I am proposing that we jointly decide that you will help me move. If I am deliberating in good faith, then I must believe that my

---

[6] Faced with a similar request, Phoebe – a character in the sitcom TV show *Friends* – responds, "Oh, I wish I could, but I don't want to." It's a laugh line because so many of us have wanted to say exactly this, but felt we cannot.

proposal might be warranted: that is, I must believe that our shared reasons might support your helping me move enough to justify our jointly deciding that you will do so.[7] But if you turn down the proposal in good faith, you must believe that you are not required by our shared reasons to accept it. This means you must believe one of two things.

First, you could think that while the decision that you will help me move would be warranted, the decision that you *won't* help me move would be warranted as well. You could think that these two shared options are on a par, both permissible but neither required. (In other terms: you might think that our shared reasons in favor of your helping me move are *sufficient*, but not *decisive*). If this is the case, however, then by rejecting my request, you are resolving the tie between these shared options in a way that clearly favors your preferences over mine. To reject my request in this tiebreaking scenario, then, you have to show a willingness to be mildly, but blatantly, selfish.

Second, you might believe my proposal to be simply unwarranted. But if I proposed this joint decision in good faith, *I* must not believe that it is unwarranted. So if you think it is unwarranted, you must either think that I misjudged the balance of our shared reasons, or you must know something I don't that counts against the proposal.

The implicature of your turning down my request, then, is that one of three things is the case: either (a) you are being selfish, (b) you think I misjudged the weight of our shared reasons, or (c) there is some reason against your accepting the request of which I was not aware. Given these options, no wonder we prefer to opt for (c) and offer an excuse for turning down the request. (a) is unattractive for obvious reasons. (b) is perhaps worse: if I misjudged our shared reasons, but not because I was unaware of some relevant matter of fact, that means I must have overestimated how much weight you attach to our shared reasons. In short, I must be mistaken about how important our relationship is to you. So the implicit message of both (a) and (b) is, "I

---

[7] To make a proposal in good faith, I don't have to believe outright that the proposal *is* warranted; I only have to *not* believe that it is *un*warranted. So, I can propose a joint decision in good faith while being uncertain about its warrant. Indeed, a central function of offers and requests is to enable us to do just this.

117

don't care about you *that* much." Even when this *is* true, we don't want to acknowledge it so openly. This is why we feel pressure to offer a reason for turning down the request of which the speaker was not aware (c): for instance, a prior plan that conflicts with fulfilling the request. The alternative is to admit that we don't care as much about the other person as they think we do.

Thus the deliberative theory predicts that turning down a request will be awkward in exactly the ways we experience it to be. I don't see how the reductive theory could offer a similar explanation. The norms of desire expression do not seem to require that one express a desire that someone ϕ only if one believes they might have sufficient reason to ϕ. Indeed, it is sometimes perfectly appropriate to express a desire for someone to do something you *know* they ought not do: "I know you have to go, but I want you to stay." If requests are just expressions of desire, it is unclear why we are so careful about making them, and so reluctant to turn them down.

The view that requests are promise solicitations thus appears too impoverished to explain the social significance of making a request. In requests, again, we have a speech act that is strikingly similar to promises, but which does not seem reducible to them. The opponent of the deliberative theory thus faces yet another mystery: if requests are not reducible to promises, then why are they so similar to them?

3.3. Requests and demands

Requests are to demands as offers are to promises: requests propose decisions regarding the addressee's actions by means of propose-and-ratify, while demands propose the same kind of decisions using propose-and-challenge.[8] This should lead us to expect that the differences between offers and promises' appropriateness conditions surveyed in §1.3 will be paralleled by requests and demands. And when we look, this is just what we find.

---

[8] For our purposes in this section, we need not distinguish between demands and commands. So, rather than repeat the cumbersome phrase "demands and commands" throughout, I will speak only of demands. But the observations I make should apply to commands as well. For more on this distinction, see the next chapter.

To review, the propose-and-challenge and propose-and-ratify methods differ in two main respects. First, they guard against different types of deliberative error: propose-and-challenge protects against failing to make joint decisions that are warranted (false negatives), while propose-and-ratify protects against making joint decisions that are unwarranted (false positives). Second, they distribute deliberative power in different ways: propose-and-challenge puts greater power in the hands of the speaker, while propose-and-ratify gives power to the addressee.

These differences allow us to make four predictions about when it is appropriate to make requests and demands, which roughly parallel the predictions we made for offers and promises. Let me review each prediction in turn.

(i) *Demands are appropriate only when the speaker justifiably takes herself to know that the demand is warranted. Conversely, only requests are appropriate when the speaker is uncertain whether the request is warranted.* Like offers, one function of requests is to air proposals that one doesn't know will be, or ought to be, accepted. My colleague is very busy, but I'd like to get her thoughts on my latest draft. I genuinely don't know whether she can or should take the time to read my draft. The appropriate thing to do is ask: "Any chance you could read my draft?"

Not all requests are made under such uncertainty. I will request rather than demand a meeting with my advisor, even when I know that he has sufficient reason to meet with me if I ask (it's his job, after all!). It is usually polite to make requests rather than demands even when one knows they ought to be accepted (see (ii) below). But even when demands would otherwise be permissible, uncertainty about warrant renders them inappropriate.

Consider a head chef at a restaurant. If the head chef knows what the sous chef should be doing, she can just demand that he do it: "Chop the onions!" But suppose she is uncertain about a particular case: she would like the sous chef to take over plating the appetizers so that she can focus on the entrée, but she doesn't know if the sous chef has time to do that while fulfilling his other more important duties, such as making the soup. In this case, it would be brash for her to simply demand that the sous chef plate the appetizers. Instead, she should ask: "Can you take over plating the appetizers for me?" This gives the sous chef an opportunity to check whether the

proposal is warranted, and say, "Sorry, I've got my hands tied with the soup," if it's not. When the warrant of a proposed decision is under question, it is typically wise to give the addressee a chance to double-check the proposal before it is accepted. This is why requests are more appropriate under conditions of uncertainty, as the deliberative theory predicts.

(ii) *Ceteris paribus, requests are more appropriate than demands, because they give the addressee more deliberative power.* It seems undeniable that requests are far more common than demands in everyday life. Genuine demands are rare occurrences, common only in contexts of formalized authority (e.g., the military or a restaurant kitchen) or moments of high drama ("Let my people go!"). In comparison, requests are innocuous creatures, perfectly at home in polite conversation.

Why would this be? As we discussed in the case of offers, it seems to be good deliberative practice to err on the side of giving more power to one's addressee. But this point has special significance for requests and demands. Since requests and demands propose joint decisions that concern the addressee's actions, it is particularly important to give the addressee power over whether these decisions are made. By making a demand, one seizes asymmetric deliberative power over another person's actions. The bar for doing so appropriately is thus understandably high. We can thus predict that there will be a particularly wide gap between the appropriateness conditions of demands and requests, since there is special reason to use propose-and-ratify when the addressee's actions are at stake. This fits our intuitive judgments: there seems to be a much bigger difference between making a demand and making a request than there is between making a promise and making an offer.

(iii) *The more likely the addressee is to reject a warranted proposal, the more appropriate it is to make that proposal using a demand rather than a request.* Sometimes we start out by making a request, but then turn to demands when the request is ignored. "Would you pass the salt, please? ... Hello, can you pass the salt? ... *Hey!* Pass the damn salt!" Parents often follow this pattern when dealing with stubborn children. "Sam, would you please put away your toys?" "No!" "Sam, *put away your toys.*"

When an addressee is not adequately doing his part in the deliberation – by, for example, ignoring or turning down a request he ought to accept – it is sometimes reasonable to respond by

taking deliberative power away from him. Knowing that Sam ought to put away his toys, I might still request rather than demand that he do so, giving him the opportunity to actively accept this decision. When he refuses, however, I may be warranted in taking away this deliberative power by making a demand. This puts the justificatory burden on Sam: if he wants to block my proposal, he will have to provide some legitimate challenge to it, rather than just saying "No!" The deliberative theory allows us to see this request-demand progression as not just an expression of increasing frustration, but as a strategic shift in the balance of deliberative power.

This power dynamic also explains why demands are better tools for holding people accountable to their moral obligations. If you see me doing something horrible, it is more appropriate for you to demand "Stop that at once!" than to meekly request, "Would you please stop?" When someone is already doing wrong, he has thereby demonstrated a propensity to disregard his obligations. Thus you should take the deliberative power into your own hands, *demanding* that he stop doing wrong rather than merely requesting that he do so.

(iv) *When there is reason to give the speaker greater power over the addressee's actions, demands can be appropriate.* Sometimes we have good reason to put someone in charge. There are head chefs because kitchens function better when a single person sets the menu and decides how to divide up the labor. Armies have hierarchies of command because they function better when some people give orders and others obey them. If we are unloading a moving van, we might appoint one person the 'leader' in order to better coordinate our actions.

The propose-and-challenge method's greatest liability is also its most important asset: the power it confers on the speaker. This power dynamic explains why it is usually inappropriate to use the propose-and-challenge method when putting forward decisions about others' actions: because by doing so, you assert asymmetric power over the choices of others. But this power dynamic can also be an advantage, because sometimes we want to put the decision-making power in a single person's hands. As I argue at length in the next chapter (especially §4.2), the situations when we stand to benefit from putting the power in one person's hands are exactly the situations in which demands and commands are intuitively appropriate. Yet another illustration

of this section's wider point: by thinking about the costs and benefits of the propose-and-challenge and propose-and-ratify methods, we can explain our commonsense judgments about when it is appropriate to make a demand, and when it is better to make a request.

4. The unity argument

The above discussions of offers, agreements, and requests were cast from a common mold. I began by pointing out the similarities between each of these speech acts and promises. When I make an offer and you accept it, the moral situation is just as if I had made a promise. When we make an agreement, the mutual obligations we thereby undertake have the same shape and significance as promissory obligations. When I accept your request, I become obligated to do what you asked in the same way as if I had promised to do so.

Then I asked a question: how do we explain these speech acts' similarities to promises? The deliberative theory offers an easy answer to this question. Promises, offers, agreements, and requests are all species of the same genus. When successful, these acts all have the same effect: they bring a new joint decision into force. This new joint decision gives promises, offers, agreements, and requests their shared pattern of normative effects. Joint decisions put their subjects under an obligation with a certain suite of normative features. When we look at these features, we find that they are the very features that are common to the obligations created by promises, offers, agreements, and requests. The other salient similarity between these acts – that all of them are invalidated by coercion and deception – is explained by the fact that coercion and deception are forms of deliberative bad faith. By subsuming promises, offers, agreements, and requests under a common framework, the deliberative theory enables us to predict and explain their deep similarities.

Life is not so easy for other views. Their best option seems to be to claim that offers, agreements, and requests are so similar to promises because they *are* promises of special kinds. If this view could be defended, it would provide an alternative explanation of these speech acts' commonalities.

122

But, as I have argued in this chapter, this reductive view is false. Offers, agreements, and requests cannot be plausibly reduced to promises. The best attempts at such reductions fail to capture essential features of each kind of act. We cannot hold that offers are conditional promises, for this fails to capture the fact that unaccepted offers expire and can be revoked (§1.2). We cannot hold that agreements are exchanges of promises, for no promise-exchange can generate the interdependent obligations characteristic of agreements (§2). We cannot hold that requests are promise solicitations, for while this view captures the normative significance of accepting requests, it fails to explain the normative impact of *making* them (§3.2). Thus the most plausible alternative account of the similarities between promises, offers, agreements and requests seems to be ultimately untenable.

This provides the materials for a second general argument for the deliberative theory. (The first was the redundancy argument presented in Ch. 2, §5). Any adequate theory of promises, offers, agreements, and requests must explain why these four kinds of act have such similar normative effects and validity conditions. The deliberative theory can explain this observation: these four acts are so similar because they all result in joint decisions. Other theories seem to offer no viable explanation of this fact. Their most plausible candidate – the view that offers, agreements, and requests reduce to promises – simply fails to capture the intuitive features of offers, agreements, and requests. And so the opponent of the deliberative theory is left with a mystery: if offers, agreements, and requests are not special kinds of promises, then why are they so similar to them? It seems that the only way to answer this question is to hold, with the deliberative theory, that these acts are all moves within joint practical deliberation. Call this *the unity argument*, as it challenges alternative theories to explain the *unity* of promises, offers, requests, and agreements.

Given what we know about joint practical deliberation, we should expect to find a plurality of distinct speech acts with very similar normative upshots. There are different ways of making joint decisions, and different contents that joint decisions can have. Putting these parameters together, we get a set of possible moves within joint practical deliberation that look

exactly like moves we recognize in ordinary language: promises, offers, agreements, and requests. The deliberative theory predicts the differences between these acts while explaining their unity. It offers a single map on which we can locate the full variety of ways we make commitments to each other. I have yet to find another theory that does the same.

Chapter 4

Demands and Commands

Consider the following cases:

SAND CASTLE: Walking along the beach, you come across me stomping on a beautiful sand castle. To your horror, you realize that the sand castle is not mine, but instead belongs to a nearby child, who is watching tearfully as I destroy her creation. Seeing this, you shout: "Stop that at once!"

PUSH-UPS: The sergeant in charge of a training camp likes to keep her soldiers on their toes. Walking through the mess hall during lunch, she picks out a soldier at random and barks: "Johnson! Drop and give me fifty!"

These two examples are both importantly similar and strikingly different.

On the one hand, your utterance in SAND CASTLE and the sergeant's utterance in PUSH-UPS share many features: both take the imperative grammatical form; both seek to get their addressee to perform an action; both are forceful in a way that other forms of interpersonal influence – say, making a request or giving advice – are not. On the other hand, these utterances also differ in crucial respects. When she tells Johnson to do fifty push-ups, the sergeant obligates him to do something he had little reason to do before; when you tell me to stop destroying the sand castle, you are merely holding me to an obligation I already had. The sergeant has the authority to tell Johnson to do push-ups because of her special role in an institution; you have the authority to tell me to stop destroying the sand castle simply because you are another human being.

Following ordinary usage, let us call your speech act in SAND CASTLE a *demand*, and the sergeant's speech act in PUSH-UPS a *command*. (Also following ordinary language, I will call commands 'orders' sometimes; please treat the two words as synonyms).

In this chapter, I will argue that demands and commands are both proposals to make joint decisions regarding the addressee's actions by means of the propose-and-challenge method.

Where demands and commands differ is in the kind of decision they propose. Demands propose what I will call *commissive decisions*: decisions to perform an action one already has decisive reason to do. Commands propose what I will call *enactive decisions*: decisions to perform one of out of multiple actions that are rationally on a par. So, the theses I will defend are:

> **The Deliberative Theory of Demands:** For S to demand that A φ *just is* for S to propose to A, by means of the propose-and-challenge method, that they make a commissive joint decision to the effect that A will φ.

> **The Deliberative Theory of Commands:** For S to command A to φ *just is* for S to propose to A, by means of the propose-and-challenge method, that they make an enactive joint decision to the effect that A will φ.

I begin the chapter by elaborating on the contrast between demands and commands (§1). Despite their important differences, I argue, the commonalities between these speech acts are deep enough to require explanation. However, the prospects for reducing one of these speech acts to the other appear dim (§1.2). Instead, I suggest, we should understand the contrast between demands and commands in terms of the distinction between commissive and enactive decisions (§2). In the following sections, I elaborate upon and defend the deliberative theory of each speech act, starting with demands (§3) and then moving to commands (§4).

1. Two kinds of authority

1.1. Commands and demands: differences and similarities

Let's take a closer look at the contrast between SAND CASTLE and PUSH-UPS.

The first and most salient difference between the two concerns the relation between the speaker's utterance and the addressee's reasons for complying with it. In PUSH-UPS, the sergeant's command creates the soldier's reasons for complying with it. Before the sergeant gives the command, Johnson (the soldier) has no reason to interrupt his lunch to spontaneously drop to the floor and do fifty push-ups. But once the sergeant gives the order, Johnson has decisive reason to drop to the floor and do fifty push-ups. The sergeant's command does not merely trigger some

independent reason Johnson had to do push-ups; her command *is* the reason why Johnson should do push-ups. If Johnson asks, "Why should I do that?" the sergeant can answer: "Because I said so!"

Compare SAND CASTLE. Here, you do not take your demand to be the reason why I ought to stop destroying the child's sand castle. Perhaps your demand provides some extra reason for me to do so, but this reason is superfluous: the fact that I am destroying a child's valued creation, taken alone, is enough to give me decisive reason to stop. Instead of creating a new reason, the point of your demand seems to be to hold me accountable to the reasons I already have. If I responded to your demand by saying, "Okay, I'll stop, but just because you said so," you could reply, "Don't stop because *I said so*, stop because *it's wrong!*"

The second major difference between commands and demands concerns the kind of authority one must have to make them. The authority to command is a rare thing: only particular people have it in particular situations. To have the authority to command, one must normally occupy a special role or office within some social institution: sergeant, teacher, coach, chef, president, conductor, parent. Typically this authority only applies within a delimited domain. For example, the head chef can order around her sous chefs in the restaurant kitchen, but cannot tell them what to do in their kitchens at home. Finally, the authority to command is fundamentally *asymmetrical*: if I have authority to command you within a domain, then you do not have authority to command me within that same domain. Due to this hierarchical nature, I will call the authority to command *hierarchical authority*.

In contrast, the authority to make the kind of demand you make in SAND CASTLE appears to be universally held by all persons. *Anyone* who came across me stomping on the child's sand castle would have the standing to demand that I stop. You do not have to occupy some special institutional role to have the authority to demand that I comply with my moral obligations; you seem to get that authority merely from your status as a fellow human being. Thus the authority to demand is *symmetrical*: if you have the authority to make demands of me, then I must have the authority to make similar demands of you (cf. Darwall 2006: 20-22). Finally,

the authority to demand does not seem to be domain-specific like the authority to command: if I am doing wrong, you can demand that I stop, no matter where I am doing this wrong or whom I am doing it to. Since the authority to demand is given universally to all morally competent persons, I will call it *moral authority*.

Given these differences, should we simply conclude that demands and commands are two separate topics, requiring separate treatments? No. Despite their differences, demands and commands also have a deep similarity that calls out for explanation.

The fundamental commonality between demands and commands is this: both speech acts purport to *close the question* of what the addressee will do. Compare two other forms of interpersonal influence, advice and threats. Though advice and threats also aim to influence their addressees' actions, they leave the ultimate decision of what to do up to the addressee. Suppose I want you to read my favorite book. I might advise you: "You should read it, I think you'll really enjoy it." Or I might threaten you: "If you don't read this book, I won't watch your favorite movie." In either case, you could sensibly reply, "I'll take that under consideration." But if I *demand* that you read my favorite book – "Read this book!" – it wouldn't make sense for you to say, "I'll take that under consideration." This response could only be a joke.

Advice and threats aim to influence their addressees' actions by providing them with some new input to their deliberation. When I give you advice, I point out facts and features of your situation that support some particular course of action. When I make a threat, I provide you with a new incentive to perform a certain action (namely, that I will carry out my threat if you don't). In both cases, I am simply providing you with another reason to take into account in your deliberation; I am still leaving the decision itself up to you.

But when I demand that you do something, or command you to do it, I am not merely trying to provide some new consideration for you to take into account when you deliberate. I am purporting to end your deliberation entirely, deciding *for you* that you will do the action I am demanding or commanding. If you continue to treat it as an open deliberative question whether to do what I have said, treating my demand or command as something to be 'taken under

128

consideration', then you have thereby *rejected* my demand or command. If you think my demand or command is legitimate, you must treat it as decided that you will do what it requires.[1]

As far as I can tell, demands and commands are the only speech acts that purport to close their addressee's deliberation in this way. This suggests that demands and commands may be two subtypes of a single overarching kind of speech act, the distinctive force of which is to decide for the addressee what she will do. If we can characterize a wider speech act genus of which demands and commands are species, then we will be able to explain their similar force. This is just what the deliberative theory does: it locates a single speech act category within which both demands and commands fall, and explains their similarities by appeal to the features of this wider category. Can other theories offer the same kind of explanation?

1.2. Extant accounts

Demands and commands are rarely considered together: the literatures on these two speech acts are largely separate from one another.[2] Thus I will look at the leading theories of commands and demands separately, and ask whether each could be extended to capture the other speech act.

1.2.1. Raz's theory of commands

The most influential contemporary theory of authority to command is that of Joseph Raz (1985; 1986; 2006; 2010). On Raz's view, the signature feature of hierarchical authority is the *surrender of judgment* it involves. I should give you hierarchical authority over my actions when, and because, I think you are better positioned to judge what I should do than I am. When this is the case, I should just do what you say instead of acting on my own judgment. Instead of assessing

---

[1] This contrast between demands and advice is noted by both Darwall (2006: 5-7) and Macnamara (2009: 97). Macnamara proposes that demands are more like threats, motivating action by threatening the sanction of blame (2009: 92). I think this view should be rejected for the same reasons why we should think demands are distinct from advice: threats aim only to provide an input to deliberation, while demands purport to close it.

[2] One exception is the work of Stephen Darwall, especially his (2013a) and (2013b). But, I argue below, it is not clear how his account of demands might be extended to explain commands' power to create new obligations.

the balance of my reasons myself, I should simply take your command as my reason for action. This is the rationale behind what Raz calls the *preemption thesis*: "The fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of some of them" (Raz 1986: 46). This fits the observation we just made: commands do not seem to offer further reasons to be taken into account in deliberation, but instead to *preempt* deliberation, replacing the addressee's judgment with the authority's directive.

To learn whether a purported authority is legitimate, then, all we have to do is ask whether we should surrender our judgment to them in this preemptive way. This leads to the centerpiece of Raz's theory, which he calls the *normal justification thesis*:

> The normal way to establish that a person has authority over another person
> involves showing that the alleged subject is likely better to comply with reasons
> which apply to him ... if he accepts the directives of the alleged authority as
> authoritatively binding and tries to follow them, rather than by trying to follow the
> reasons which apply to him directly (Raz 1986: 53).[3]

In other words, you have authority to command me within a certain domain just in case I am likely to better comply with my reasons for action in that domain by obeying your commands than I am by deciding for myself what to do.

If the purpose of commands is to help their addressees better act on their own reasons for action, then a command had better be based upon the reasons it is supposed to preempt. Thus the normal justification thesis leads naturally to the *dependence thesis*: "All authoritative directives should be based on reasons which already independently apply to the subjects of the directives and are relevant to their action in the circumstances covered by the directive" (47).

These three theses comprise Raz's account of both the force of commands and the conditions under which one has the authority to make them. The force of commands, we learn

---

[3] What I am calling a 'command,' Raz calls a 'directive.' The difference is only terminological.

130

from the preemption thesis, is to preempt judgment: to give the addressee a decisive reason for action that stands in for, and replaces, her judgment of the balance of her other reasons for action. For a command to be authoritative, we learn from the dependence and normal justification theses, it must be based on the addressee's reasons for action, and the addressee must be more likely to conform to those reasons by obeying the command than she would be by deciding herself.

I do not wish to evaluate Raz's account of commands here (but see §4.4 below). Instead, my question is: can Raz's theory of commands explain demands as well?

Begin with the force of demands. Raz's preemption thesis captures our observation that demands purport to close deliberation. However, the way in which Raz thinks commands preempt deliberation does not fit demands as naturally. For Raz holds that commands preempt deliberation by providing a reason that is "meant to replace the reasons on which it depends" (42). So, when the head chef orders the sous chef to chop the onions, the sous chef is now supposed to chop the onions not for the reasons he had to do beforehand (e.g. that he needs chopped onions for the soup), but instead for the reason *that the head chef told him to*. While this rings true for commands, it sounds false for demands. As we have already observed, when you demand that I stop stomping on the child's sand castle, you don't take your demand to be the reason that should motivate me to stop. You think I should be motivated by the facts that make my stomping wrong – e.g., the sadness I am causing the child. If I say, "I'll stop, but only because you said so," I am missing the point of your demand. So, while your demand purports to close my deliberation, it does not seem to do so by providing a reason that replaces my other reasons for complying with it.[4] The preemption thesis appears false when applied to demands.

Let's now turn to the conditions for authority to demand. Raz's criterion for hierarchical authority, the normal justification thesis, does not seem to fit the facts about moral authority.

---

[4] This is not to say that demands are motivationally inert. In making a demand, one typically hopes that the demand will help to motivate one's addressee to do the right thing. But the way in which demands motivate does not seem to be by *replacing* the addressee's other reasons for action. Instead, demands seem to motivate either by drawing the addressee's attention to the reasons he already had, or by supplementing those reasons with the further motivation provided by the demand. Thanks to Kieran Setiya for discussion here.

Depending on how we interpret it, the normal justification thesis either draws the bounds of moral authority too broadly or too narrowly.

First, suppose we interpret the normal justification thesis as applying to individual actions on a case-by-case basis. Then it seems that your demand in SAND CASTLE meets Raz's criteria, since I am clearly more likely to conform to my reasons by complying with your demand than by following my own judgment. If a person is obligated to ϕ, but is inclined not to do so, she will better conform to her reasons by complying with a demand that she ϕ than by deciding herself what to do (and therefore not ϕing). Thus the normal justification thesis, on this interpretation, rightly predicts that we can demand that people comply with their moral obligations.

The trouble is that the same rationale applies to any action that one ought to do at all, not just for moral reasons. If I ought to ϕ, but am inclined not to do so, I will better conform to my reasons by complying with a demand that I ϕ than by deciding myself what to do (and therefore not ϕing). Thus our first interpretation of the normal justification thesis entails that whenever one ought to do something, others can legitimately demand that one do it. But this is starkly at odds with common sense. I ought to floss every night, watch my sugar intake, and work on this dissertation rather than watch Netflix – but nobody can *demand* of me that I do these things. If you see me chowing down on sugary cookies and say, "Stop that at once!" I could legitimately reply: "Mind your own business!"

The right fix seems to be to reinterpret the normal justification thesis as applying to whole domains of action. Instead of asking whether one will better comply with one's reasons for action by obeying a particular command, we should ask whether one will better comply with one's reasons for action within a certain domain by obeying a person's commands *in general*. The head chef has the authority to command the sous chef in the restaurant kitchen because the sous chef is more likely to *generally* comply with his reasons for action by obeying her commands rather than deciding what to do by himself.[5]

---

[5] Raz seems to favor this second interpretation (see Raz 1986: 60-62).

On this second interpretation, the normal justification thesis does not imply that everyone has authority to demand that I floss. For, in the domain of my personal hygiene, I am generally in the best position to judge what I ought to do. But now the normal justification thesis does not generate *enough* authority to demand. Take your demand in SAND CASTLE. Suppose that, with the notable exception of my current behavior, I am generally a paragon of moral virtue. I am quite good at judging the balance of my moral reasons and acting in accordance with them. Suppose also that your judgment on moral matters is fairly unreliable. Then it seems possible that, despite my wrongdoing in this particular case, I am more likely to conform to my moral reasons in general by deciding for myself what to do than by following whatever demands you make of me. Thus our second version of the normal justification thesis would predict that you do not have the authority to demand that I stop destroying the child's sand castle. But this is clearly incorrect. No matter who is a more reliable judge of my moral obligations in general, the fact is that in *this* case I am doing wrong, and that alone is enough to give you the authority to demand that I stop. Intuitively, *anyone* has standing to demand that I stop doing wrong, regardless of whether their moral judgment is more reliable than mine. Thus the normal justification thesis, on its second (and more plausible) interpretation, draws the bounds of moral authority too narrowly.

I conclude that Raz's theory of commands cannot be plausibly extended to demands. It mispredicts both the reasons that ought to motivate compliance with demands and the extension of our standing to make them.

1.2.2. Darwall's theory of demands

Perhaps we should instead start with demands, and try to explain commands in terms of them. Thus let us consider Stephen Darwall's influential theory of demands (2006).

On Darwall's view, demands and moral obligation are conceptually intertwined: an action is morally obligatory just in case we can demand it of one another, and a demand is authoritative just in case it expresses a moral obligation (Darwall 2006: 101). Demands and moral obligation are part of a wider circle of phenomena that involve what Darwall calls "the

second-person standpoint," which is "the perspective you and I take up when we make and acknowledge claims on one another's conduct and will" (3). Whenever we make a demand, we take up this standpoint and thereby commit ourselves to certain interpersonal standards of justification. When I make a demand of you, I must recognize your standing to make similar demands of me (117), and I must take my demand to be one that you can recognize as legitimate from your own perspective and make of yourself in turn (112).

This theory provides an elegant framework within which to understand demands' force and justification. (One that, I argue briefly in §3.2, coheres nicely with the deliberative theory). But taken alone, this framework will not give us a satisfactory account of commands.

As Darwall argues, a demand is only authoritative if its addressee is obligated to do the action demanded. However, commands can be authoritative even when their addressees were under no previous obligation to do the action commanded. The sergeant's command in PUSH-UPS is authoritative even though Johnson had no prior obligation to do fifty push-ups at that moment. To extend Darwall's theory to commands, then, we might hypothesize that commands work in two steps: first, they create a new obligation, and second, they demand that the addressee comply with it. Thus when the sergeant commands Johnson to do fifty push-ups, her utterance both creates an obligation for Johnson to do the push-ups *and* holds him accountable for complying with it. It is as if she had first said, "I hereby make it the case that you, Johnson, are obligated to do fifty push-ups now," and then demanded: "So, *drop and give me fifty!*" Commands are thus demands that hold their addressees accountable to a specific kind of obligation: namely, an obligation to do what the speaker says.

This account captures the similarities between commands and demands by building demands into commands as a constituent part. However, as an account of commands, this proposal is deeply unsatisfactory. It presupposes what an account of commands is supposed to explain: namely, commands' power to create new reasons for their addressees. The first step in the two-step process outlined in the last paragraph is left completely unexplained. How can the sergeant create an obligation for Johnson by simply declaring that it exists? Why does she have

the authority to put Johnson under obligations in this way, while Johnson does not have the authority to do the same to her? What is the basis of the reason that the sergeant's command gives to Johnson? These are the questions we want our theory of commands to address; Darwall's account of demands, taken alone, gets us no closer to answering them.[6]

To be clear, the arguments of this section are not criticisms of Raz's theory of commands or Darwall's theory of demands in themselves. Neither theorist intended for their account to be extended to the other speech act in the way we just attempted. My point is not that Raz or Darwall's theories fail on their own terms, but that they cannot provide what we are looking for here: a unified account of commands and demands that explains the similarities between these speech acts while respecting their differences. To find such an account, we have to look elsewhere.

2. Two kinds of decision

Forget about demands and commands for a minute. Consider these examples instead:

GYM: I know I ought to go to the gym. Exercise not only promotes my long-term health, it also makes me feel happier and more energetic in the short term. Though it is tempting to sit on the couch watching TV, there's no question in my mind that going to the gym is what I ought to do. So I decide to go to the gym.

LUNCH: I'm deciding where to get lunch. There are two nearby restaurants, Darwin's and Clover. Thinking about it, I don't see any strong reason to choose one over the other; the two are on a par, roughly similar in price, quality, atmosphere, distance, etc. But I have to choose *somewhere*, so I decide to go to Darwin's.

These cases illustrate two different roles decisions play in our lives. Sometimes, as in GYM, we make decisions in order to commit ourselves to doing what we have most reason to do. But other

---

[6] Darwall seems to acknowledge this when, rather than claiming that his theory explains hierarchical authority, he says only that it provides the framework within which such explanations must be given: "the only justification [of hierarchical authority] that can succeed is one that proceeds from within the second-person standpoint, beginning with the assumption that we all share a common basic authority ... and proceeding from there to consider what differential claims to authority anyone could sensibly accept" (2013b: 167).

times, as in LUNCH, our reasons don't supply clear guidance about what to do. Sometimes we have two or more options that are on a par, our reasons supporting neither decisively over the other. Like Buridan's ass, who famously starved because he could not choose between two equidistant bales of hay, our judgments of what we ought to do are not sufficient to determine our behavior in these cases. Decisions step in to fill the gap: by making a decision, we commit ourselves to one of the equipoised options and avoid the fate of Buridan's ass.

Call decisions of the first kind *commissive decisions*, as they *commit* us to acting on reasons we already have. Call decisions of the second kind *enactive decisions*, as they *enact* a determination of action that outstrips the balance of reasons.

Sometimes, among the options available to a person, one is clearly the best choice. We can describe this fact in multiple ways: the person has *decisive reason* to perform this option; she is *rationally required* to perform this option and *rationally forbidden* to perform the alternatives; she *ought* to perform this option and ought not perform any other. I will treat these formulations as equivalent. When a person decides on an option because she judges that her reasons decisively support it, she makes a commissive decision. Thus my decision in GYM is commissive: I decide to go to the gym because I judge that I am rationally required to do so.

If there is no single option that a person has decisive reason to perform, then there must be multiple options, each of which she has sufficient reason to perform.[7] Though there are almost always options that we clearly ought *not* perform, there are many cases where no single option stands out as best. Often two or more of the options available to an agent are on a par.[8] Again, philosophers have multiple ways of describing this situation: we can say that an agent has *sufficient reason* to perform each of the options; that each of the options is *rationally permissible*, but none are rationally *required*; that she ought to perform *one* of the options, but none of them is such that she

---

[7] I assume that there are no *rational dilemmas*: cases where a person has no options that has sufficient reason to perform (i.e., no options that are rationally permissible). If there are rational dilemmas, then I suspect we should classify decisions in such cases as enactive.

[8] I take the phrase "on a par" from Chang (2002). Chang argues, persuasively I think, that two options can be on a par without being exactly tied: i.e., without it being the case that adding a very small reason for performing one would 'break the tie' and make it decisively favored over the other. Perhaps this is because there is vagueness in comparisons between reasons, or perhaps, as Chang claims, it is because some of our reasons are incommensurable.

ought to perform *it*. I will treat these formulations as equivalent. When a person decides on one of multiple options that she judges to be on a par in this way, she makes an enactive decision. Thus my decision in LUNCH is enactive: I decide on Darwin's, though I judge that I had just as much reason to go to Clover instead.[9]

Just as we can distinguish between commissive and enactive individual decisions, we can also distinguish between commissive and enactive *joint* decisions. Suppose that among the shared options available to us, there is one option that our shared reasons decisively support over all the others. The joint decision that we perform this option is not only warranted by our shared reasons; it is *required* by them.[10] When we make a joint decision because we take it to be decisively supported in this way, we make a commissive joint decision. Our joint decision simply commits us to acting as our shared reasons require.

Suppose instead that there is no shared option that is required by our shared reasons, but instead multiple shared options that are permitted by them. That is, there are multiple joint decisions we could make, each of which is warranted by our shared reasons. When we jointly decide on one out of multiple shared options that we judge to be on a par in this way, we make an enactive joint decision. Our joint decision determines what we will do in a way that goes beyond our shared reasons.

I propose that we understand the distinction between demands and commands in terms of the distinction between commissive and enactive joint decisions. Demands propose commissive joint decisions; commands propose enactive joint decisions.

The central difference between demands and commands, I have argued, concerns the reasons they offer their addressees. Demands tell their addressees to do something they were already required to do, not because it is being demanded but because of the decisive reasons they already have to do so. Commissive decisions do the same. When I decide to go to the gym, I am

---

[9] For an account that emphasizes the commissive function of decisions, see Holton (2004); for an account that emphasizes their enactive function, see Watson (2003).

[10] I say more about what it takes for our shared reasons to require an action in §3.1 below.

telling myself to do something I was already required to do, not because I decided to do it but because of the decisive reasons I already have to do so. In contrast, commands determine their addressees' actions in a way that goes beyond their prior reasons, telling them to do something simply because it is what was commanded. Enactive decisions do the same. When I decide to go to Darwin's instead of Clover for lunch, I determine my own actions in a way that goes beyond my prior reasons, telling myself to go to Darwin's simply because it is what I decided to do.

The feature that ties demands and commands together is their force: the way in which, unlike advice or threats, they purport to close the question of what their addressees will do. Demands and commands have this force, I suggest, because they put forward joint decisions by the propose-and-challenge method. When you propose the joint decision that I will φ using the propose-and-challenge method, then unless I can raise a convincing objection to your proposal, it is thereby decided between us that I will φ. Once I accept your proposal as warranted, there is no room for further deliberation on my part about whether to φ. This contrasts with advice and threats, which do not propose decisions at all, but instead just put forward considerations to be taken into account in deliberation. Other proposals to make joint decisions (promises, requests, etc.) have the same exclusionary effect on deliberation once accepted, as I have pointed out before (see especially Ch. 2, §1.3). But the effect is more dramatic with demands and commands, because they enable one person to impose this exclusionary effect on another's deliberation.

The distinction between commissive and enactive joint decisions applies to our other speech acts as well. I might promise you that I won't read your diary, though I was already obligated not to do so: a commissive joint decision. Or, I might promise to read your draft by Tuesday, when I could just as well have promised to read it by Thursday, or not at all: an enactive joint decision. But this distinction has particular significance for proposals to make joint decisions concerning others' actions by means of the propose-and-challenge method. For, as I have argued (Ch. 3, §3.3), there is a strong presumption against using propose-and-challenge to propose joint decisions concerning others' actions. By doing so, you take control over your

138

addressee's actions out of her hands and into your own. Since people should generally be given control over their own actions, this seizure of deliberative power requires justification.

Ordinary language distinguishes between demands and commands, I suggest, because they represent two very answers to this call for justification. When our shared reasons require you to φ, then proposing that you φ by propose-and-challenge is easier to justify, since I am not so much imposing my will on yours as I am holding you to the reasons you already have. This is how demands are justified: by the strength of the reasons behind them. But if our shared reasons leave open the question of whether you will φ or ψ, then by using the propose-and-challenge method to propose that you φ rather than ψ, I do seem to be imposing my will on yours. Your only reason for φing rather than ψing is that I told you to do so. Out of all of the different types of deliberative proposals, proposing an enactive joint decision regarding someone else's actions by propose-and-challenge – that is, making a command – appears to be the hardest to justify. This, I will suggest, is why we set up special institutions in which some people are explicitly granted the power to do so.

The remainder of this chapter will be spent defending these analyses, starting with demands (§3) and then turning to commands (§4).

3. Demands as proposals to make commissive joint decisions

Here, again, is our proposed analysis of demands:

The Deliberative Theory of Demands: For S to demand that A φ *just is* for S

to propose to A, by means of the propose-and-challenge method, that they make a

commissive joint decision to the effect that A will φ.

I just argued that this analysis captures both the deliberation-closing force of demands and the fact that demands, unlike commands, aim not to provide new reasons to their addressees but instead to hold them to acting on the reasons they already have. In this section I will focus on the

conditions under which demands are warranted, arguing that the deliberative theory predicts our intuitive judgments about when we have standing to make demands of one another.[11]

3.1. Giving decisive reason

There is a strong presumption against employing the propose-and-challenge method for joint decisions regarding others' actions. Demands override this presumption by appeal to the strength of the shared reasons in support of the action demanded. For a demand to be warranted, the shared reasons must *decisively support* the joint decision it proposes. This is built into the idea that demands propose commissive joint decisions. Our first task is thus to get clearer on what it means for a joint decision to be decisively supported by our shared reasons.

Way back in Chapter 1 (§2), we discussed the conditions for a joint decision to be warranted. There we were interested in what it took for a joint decision to be *permissible*; here, we are interested in what it takes for a joint decision to be *required*. But permissibility and requirement are interdefinable: an action is required just in case no other action is permissible; an action is permissible just in case one is not required *not* to perform it. So, we can apply the lessons of our earlier discussion here.

We are looking for the joint analogue of an individual's having decisive reason to do something. So, we can begin by noting that I have decisive reason to ϕ only if my reasons support my ϕing determinately more than they support any other option. This motivates a corollary of

---

[11] The normative status I am interested in when I ask whether a demand or command is warranted is also called *fittingness*. A speech act or attitude is fitting in this sense just in case it meets its own internal standard of justification: an assertion that p is fitting only if you have adequate evidence for p; it is fitting to admire a person only if she is admirable. Crucially, a response can be fitting even if there are compelling practical reasons to avoid it. I have adequate evidence for plenty of assertions that I still should not make, since they would be tactless or irrelevant. Similarly, it is fitting in our sense for a slave to demand her freedom from her master, but it is all too likely that the slave should *not* make this demand, since doing so would only bring her more abuse.

It is an important question what exactly the normative significance of these standards of fittingness is. (For discussion, see Rabinowicz and Rønnow-Rasmussen 2004). Without taking a position on that debate, however, we can recognize the difference between a response's being fitting and its being advisable all-things-considered. When I talk about commands and demands being warranted, I am interested in the former status, not the latter. For stylistic reasons, however, I will mostly avoid the term 'fitting'. Instead, I will say that one *has standing to* demand or command, that a demand or command is *warranted* or *legitimate*, or that one has the *authority* to make a demand or command. As I will use them, all of these phrases mean the same thing: that the demand or command is fitting.

140

SUFFICIENT SHARED REASON for the decisive case: a joint decision that you will φ is required *only if* our shared reasons support your φing determinately more than they support any other option.

As before, however, this condition is not enough. Since our shared reasons are only a subset of my individual reasons, it could be that our shared reasons most support my φing when my individual reasons in total most support my *not* φing. For example, if you and I are running buddies, then our shared reasons might most support my going on a run with you, while in fact I ought to be working on my dissertation instead. If you were to demand that I go running with you, I could reasonably object on the grounds that I ought to be doing something else. For a joint decision to be required, then, it's not enough for it to just be best by the lights of our shared reasons; it must also be the case that I ought, all things considered, to act in accordance with it. This was the point that motivated JOINT-INDIVIDUAL COHERENCE. The corollary here is: a joint decision that you will φ is required *only if*, conditional on the joint decision's having been made, you have decisive reason, all things considered, to φ.

In Chapter 1, we ended the discussion here, leaving it as an open question whether satisfying both SUFFICIENT SHARED REASON and JOINT-INDIVIDUAL COHERENCE is enough to make a joint decision warranted. Now, however, I want to argue that the fact that a joint decision satisfies the analogues of these conditions we just stated is not enough to make it the case that it is required by our shared reasons. To see why, consider the following example:

> **CONFERENCE:** My friend Jane is attending a conference in my area of interest at her university. If I go, then I'll get to see Jane. But the conference is several hours' drive away, and another friend is giving a talk at my home university that same weekend. However, the conference also presents a unique opportunity to meet new people and hear new ideas in my area of research. Taking all of this into account, I ought to go to the conference.

It is both the case that Jane and my shared reasons most favor my going to the conference and that, all things considered, I ought to go. And yet it seems clear that Jane could not legitimately *demand* that I come to the conference.

Why? Because our shared reasons are not enough *on their own* to make it the case that I ought to go. Though we do have a significant shared reason in favor of my going to the conference (that doing so will allow us to spend time together), this reason alone is not enough to outweigh all of my reasons *against* going to the conference, such as the fact that it conflicts with my other friend's talk. These considerations are only outweighed when I take into account reasons Jane and I do not share, such as the fact that attending will enrich my research. And so, if Jane were to propose to jointly decide that I will go to the conference, this would be asking me to make the right decision, but for the wrong reasons.

For our shared reasons to require the joint decision that I φ, then, it is not enough for it to be the case both that I ought to φ, all things considered, and that my φing is most supported by our shared reasons. Instead, I want to suggest, it must be that I ought to φ *because of* our shared reasons. In other words, our shared reasons must *give me decisive reason* to φ:

GIVING DECISIVE REASON:

For any persons X, Y, Z, et al., and action φ,

(i)   The joint decision that X will φ is required by the shared reasons of X, Y, Z, et al. *iff* those shared reasons give X decisive reason to φ.

(ii) The reasons X shares with Y, Z, et al. *give X decisive reason to φ iff*$_{Def}$ these reasons, given only the weight they have as shared reasons, weigh in favor of X's φing determinately more than all of X's reasons against φing weigh against it.

The relation of giving decisive reason defined here is analogous to the relation an action has to a person's individual reasons for action when she ought to do it. My individual reasons require me to φ just in case the reasons I have in favor of φing are enough to decisively outweigh all of my reasons against φing. Similarly, our shared reasons require me to φ just in case the shared reasons I have in favor of φing are enough to decisively outweigh all of my reasons against φing: that is, just in case our shared reasons *give me decisive reason* to φ. This is the condition that fails to obtain in

142

CONFERENCE; and, as we shall now see, it is the condition that is met in the cases where we intuitively take demands to be warranted.[12]

3.2. Who has standing to demand?

Our intuitions about standing to demand divide cases into three categories.

In what I call *everybody's business cases*, everyone has standing to demand that I perform some action. SAND CASTLE is a case of this kind: anyone who came across me destroying this poor child's sand castle could legitimately demand that I stop. Any action that is sufficiently horrible falls in this category: anyone can demand that I not deface the Mona Lisa, punch a stranger in the face, dump poison into the Pacific, or yell 'Fire!' in a crowded theatre.

On the other end of the spectrum are what I call *nobody's business cases*: cases where I ought to perform some action, but nobody has the standing to demand it of me. (I mentioned these cases briefly in §1.2.1). I ought to go to the gym, read *Anna Karenina*, and learn how to cook risotto, but nobody can reasonably demand that I do these things (cf. Baier 1966: 221; Gibbard 1990: 41). You can certainly *advise* me to read *Anna Karenina*, but to *demand* that I do so would be unreasonable (not to mention strange). This does not mean that no one could *ever* demand that I perform these actions: if I am being sufficiently self-destructive, then my loved ones might legitimately demand that I take better care of myself – and if that means going to the gym, then so be it. But it is overwhelmingly plausible that there are *some* cases where I ought to do some action that will make me healthier or happier, but nobody can demand that I do it.

I think there is also an intermediate kind of case: cases in which somebody has the standing to demand that I φ, but others do not have this standing. Call these *somebody's business cases*. Here's one such case, adapted from an example in Chapter 2 (§2.2):

---

[12] Even when the condition of GIVING DECISIVE REASON is met, it may still be more socially appropriate to make a request rather than a demand. For instance, if the speaker is uncertain whether the shared reasons give her addressee decisive reason to φ, then even if they do, it is more appropriate for the speaker to request that the addressee φ rather than demanding that she do so. I elaborate on these differences in social appropriateness at length in Ch. 3, §3.3.

DISHES: My partner Dee Dee has done much more than her share of cleaning in our apartment lately, and our sink is full of dirty dishes. She's asked me to wash the dishes several times and I have ignored her requests. Fed up, Dee Dee demands: "Brendan, wash the dishes!"

Compare our neighbor George. George has overheard Dee Dee griping over the phone about my laziness in the kitchen, and so has all of the information he needs to judge that I ought to wash the dishes. While visiting, George turns to me and demands: "Brendan, wash the dishes!"

Intuitively, Dee Dee's demand is perfectly legitimate, while George's demand is out of place. George is in a position to advise me to wash the dishes, but to demand that I do so is to overstep his authority. Whether I wash the dishes or not is none of his business.

Of course, some of the obligations I owe to Dee Dee *are* George's business: if George saw me physically assaulting Dee Dee, he (and anyone else) could legitimately demand that I stop. The point is just that there are *some* situations in which I ought to do something, and someone can demand that I do so, but nobody else has standing to make that demand. The members of my choir can demand that I come to our rehearsal, while others, intuitively, cannot. My department colleagues have the standing to demand that I do my share of committee work; but intuitively, no one else has the standing to demand this of me.

What determines how we sort cases into these three categories? The first answer that comes to mind is what I call *the interest account*: X has standing to demand that Y φ just in case X's interests would be sufficiently harmed by Y's failing to φ. This account is most appealing when we consider nobody's business and somebody's business cases. Nobody has standing to demand that I go to the gym because nobody else's interests would be significantly harmed by my failing to do so: only my interests are at stake. Dee Dee has standing to demand that I wash the dishes because her interests are harmed by my failing to do so; our neighbor George's interests, in contrast, do not significantly depend on whether I wash the dishes. The interest account explains these cases easily.

However, the interest account has a difficult time making sense of everybody's business cases. If everybody has standing to demand that I φ, then according to the interest account, this must be because *everybody's* interests would be harmed by my failing to φ. This may be a correct description of some cases: perhaps dumping poison in the ocean harms everyone's interests, since everyone has an interest in sustaining the natural environment. But consider SAND CASTLE. The interest account must say that you have standing to demand that I not destroy the child's sand castle because *your* interests are harmed by my doing so. First, it is not obvious that this is the case: are the interests of *every person* harmed every time a significant wrong is done? But even if we grant this strong claim, the interest account seems to give the wrong *explanation* of your standing to demand. You can legitimately demand that I stop destroying the child's sand castle because I am harming *the child*, not because I am harming *you*.[13]

This problem with the interest account motivates the alternative view offered by the deliberative theory. The idea is to shift our focus from *interests* to *reasons*. Though the *interests* underlying my obligation not to destroy the child's sand castle are the child's alone, the *reasons why* I should not destroy the sand castle are in an important sense *everyone's* reasons. I shouldn't destroy the sand castle because doing so causes the child pain, and everyone has just as much reason to care about the child's pain as I do. This is why everyone has standing to demand that I not destroy the sand castle: because the reasons why I shouldn't do so are *shared* by all persons.

Extending this idea, we can predict when everybody's business cases will arise: everybody has standing to demand that I φ when the reasons that I share with all morally competent persons give me decisive reason to φ. Thus everybody's business cases will arise when we ought to do something due to general, universal moral reasons that apply to us just by virtue of being persons. This seems to fit our intuitive judgments: anyone can demand that I not harm innocent

---

[13] Hart (1955: 180-181) also criticizes the interest account, though on a different basis.

people, that I not dump poison in the ocean, that I not deface objects of beauty – and these are just the kinds of action that are supported by the reasons all persons share.[14]

The deliberative theory predicts that somebody's business cases will arise when the reasons why we ought to do something are shared with some people but not others. In DISHES, Dee Dee has standing to demand that I wash the dishes because she shares the reasons that support my doing so. The fact that washing the dishes will make our apartment tidier is a reason for both Dee Dee and me to prefer that I wash the dishes. Since this shared reason is strong enough to outweigh the reasons I have against washing the dishes (e.g., that it is a tedious task), it gives me decisive reason to do so. In contrast, our neighbor George has much less reason to care about the cleanliness of our apartment. George might have *some* reason to prefer that our apartment is clean – say, out of concern for our welfare. But his reason to prefer this is much weaker than mine. Applying the UNANIMITY constraint on shared reasons I defended in Chapter 1 (§1.2), we can infer that our shared reason in favor of my washing the dishes will be no stronger than George's reason to prefer that I do so. So, the reasons I share *with George* in favor of my washing the dishes will be fairly weak: too weak to outweigh the reasons I have against doing so. Thus the deliberative theory correctly predicts that Dee Dee has the standing to demand that I wash the dishes, while George does not.

Finally, the deliberative theory predicts that nobody's business cases will arise when the reasons why we ought to do something are not shared to a significant extent with anyone else. For example, I have much stronger reason to care about my physical health and career success than anyone else does. So, actions I ought to perform on the basis of these self-regarding reasons – going to the gym, working on this dissertation – are supported by reasons that are in an

---

[14] Some theories of reasons for action – e.g., rational egoism, or 'internalist' views that hold that an agent's reasons depend closely on her desires – may deny that there are *any* reasons that all persons share. Adherents of these theories might deny, for instance, that everyone has reason to prefer that I not harm an innocent stranger: because not everyone's welfare depends on the stranger's (egoism), or because some people simply do not care about the well-being of strangers (internalism). These claims are compatible with the deliberative theory in principle. But the deliberative theory does entail that, if there are no universally shared reasons, then there are no cases in which everyone has standing to demand that one perform an action. So, the deliberative theory is committed to denying 'hybrid' views that are internalist about reasons for action but externalist about moral demands (i.e., allowing that one might legitimately demand something that one's addressee has no reason to do). Thanks to Kieran Setiya here.

important sense shared by no one else. Of course, others have *some* reason to care about my health and career success, but their reasons will be much weaker than mine. Applying UNANIMITY again, our shared reasons in favor of my going to the gym can be no stronger than the relevant other person's reason to prefer that I do so. As a result, the reasons I share with other people will usually not be weighty enough to outweigh the reasons I have against doing these actions — the time required to go to the gym, the difficulty of writing philosophy. This is why, though I ought to go to the gym and work on this dissertation, nobody else can demand that I do so: the reasons at stake are mine and mine alone.

By thinking about what reasons we share, then, we can understand when and why others have standing to make demands of us. In its emphasis on shared reasons as the basis of demands, the deliberative theory resonates with Stephen Darwall's account of demands (2006; see §1.2.2 above). One of Darwall's central theses is what he calls 'Pufendorf's Point': "In holding people responsible, we are committed to the assumption that they can hold themselves responsible by self-addressed demands from a perspective that we and they share" (112).[15] As I understand it, Pufendorf's Point is vindicated by the deliberative theory. The "perspective that we and they share" is just the perspective we take up in joint practical deliberation: the perspective of our shared reasons. When you demand that I φ, you are committed to thinking that the reasons we share within this perspective are enough, on their own, to outweigh all of my reasons against φing (by GIVING DECISIVE REASON). Thus you must think that I can require *myself* to φ on the basis of our shared reasons. Thus we arrive at Darwall's insight: we have standing to make a demand only when our addressee can make the very same demand of herself from the deliberative perspective that we share. I have argued that appreciating this fact enables us to predict and explain the contours of our standing to make demands of each other.[16]

---

[15] See also Wallace (2013): "The very considerations that give me reason to care about doing the right thing ... equally ground corresponding claims and expectations on the part of those who are affected by what I do" (30).

[16] As Darwall's work has shown, demands are closely tied up with the concept of obligation. This suggests that the deliberative theory of demands might be applied to better understand the notion of obligation. In other work I explore this possibility, arguing that the normative status of obligation can and should be analyzed in terms of shared reasons (de Kenessey ms). On the view I defend there, for X to be obligated to φ *just is* for the reasons that X shares

4. Commands as proposals to make enactive joint decisions

We turn to the deliberative theory of commands:

The Deliberative Theory of Commands: For S to command A to φ *just is*

for S to propose to A, by means of the propose-and-challenge method, that they

make an enactive joint decision to the effect that A will φ.

I will start by showing how the deliberative theory captures commands' reason-giving force

(§4.1), and then turn to the question of what it takes to have the authority to command (§4.2, 4.3).

I conclude by comparing the deliberative theory to Joseph Raz's account of hierarchical

authority (§4.4).

4.1. The normative force of commands

When the sergeant commands Johnson to drop and give her fifty, Johnson gains a strong

reason to do push-ups that he did not have before. When the head chef orders the sous chef to

chop the onions, the sous chef is now obligated to chop the onions, when before it would have

been fine to carry on stirring the soup instead. While demands hold people to the obligations they

already had, commands create obligations where none existed. How do they do this?

On the deliberative theory, commands create new obligations by bringing new joint

decisions into force. A basic premise of the deliberative theory (defended in Ch. 2, §1.1) is that we

have strong *pro tanto* reason to abide by our standing joint decisions. Thus, when the sergeant

commands Johnson to do fifty push-ups, this brings into force a joint decision to the effect that he

will do so. Since Johnson has reason to act in accordance with his standing joint decisions, he

now has reason to get on the floor and start doing push-ups.

Demands, too, bring new joint decisions into force, and thereby create new reasons for

their addressees. However, because demands propose commissive joint decisions, the reasons

---

with some other person(s) to give X decisive reason to φ. Though I cannot defend this proposal here, it illustrates how the theory of joint practical deliberation offered in Chapter 1 might be productively applied to phenomena beyond the speech acts that are the subject of this dissertation.

they create are almost always superfluous. In SAND CASTLE, I have decisive reason to stop stomping on the child's sand castle before you demand that I do. So, though your demand gives me an additional reason to stop, this makes no difference to what I ought to do.

Commands' reason-giving force is more salient because they propose enactive joint decisions. A command picks out one among multiple shared options that are on a par and proposes to jointly decide on that option. When this joint decision comes into force, it tips the balance, giving the addressee decisive reason to do the commanded action rather than the previously equipoised alternatives. Before the sergeant made her command, Johnson was permitted to either continue eating lunch or to do push-ups. The two actions were on a par as far as their shared reasons were concerned (even if Johnson would prefer to keep eating). But once the sergeant commands Johnson to do fifty push-ups, the enactive joint decision her command brings into force tips the balance. Now, Johnson is required to do push-ups, and no longer permitted to continue eating his lunch instead. Such is the power of an enactive joint decision.

The deliberative theory explains not only the fact that commands create reasons, but also the pattern of reasons they create. The obligation to obey a command has a set of features that should be familiar from our discussions of promises, offers, agreements, and requests.

First, the reason commands give is strong but overridable. A legitimate command usually obligates its addressee to obey, but if unusual circumstances arise, it can be permissible to disobey it. For example, the sous chef should not obey the head chef's command to chop the onions if she discovers that the onions are infused with poison. This is as we should expect: we have a strong reason to abide by our joint decisions, but this reason can be outweighed by other considerations.

Second, the obligation created by commands is *directed*: the addressee owes it to the speaker in particular to obey them. If Johnson refuses to do push-ups, the sergeant has special standing to complain. If the sous chef leaves the onions unchopped, he wrongs the head chef by doing so.[17] The deliberative theory predicts this as well. My obligation to abide by our joint

---

[17] This feature of commands is emphasized by Darwall (2013a), who objects to Raz's account of authority on the basis that it cannot explain how commands generate directed obligations.

decision is owed to you in particular: you have a special stake in our joint decision-making, and you have the authority to reiterate any joint decision standing between us by demanding that I comply with it (see Ch. 2, §1.2).

Third, commands require *uptake*. I cannot issue a command to an empty room and hope to bind you by doing so. Commands do not need to be explicitly accepted by their addressees like offers and requests do. Instead, like promises, all they require is that their addressees recognize them and do not raise any successful challenges to them.[18] These are just the conditions required for a joint decision to come into effect by means of the propose-and-challenge method.

Fourth, as we have noted, commands close deliberative questions. Once the head chef commands the sous chef to chop the onions, he should not treat the question of whether to chop the onions as open for further deliberation. It would be inappropriate for him to seriously consider stirring the soup instead. He should take it as decided that he will chop the onions, excluding other options from his subsequent deliberation. This is just what the deliberative theory predicts: these exclusionary effects are the characteristic upshot of making a decision, whether that decision is individually or jointly made.

Fifth, the speaker has the power to release the addressee from a command after it comes into force. Say that, as the sous chef picks up his knife to start chopping onions, the head chef says, "Actually, Cora can chop the onions instead – you can go back to what you were doing." The head chef has thereby *released* the sous chef from his obligation to chop the onions. On the deliberative theory, this power of release works by *retracting* the joint decision the initial command brought into force. Both the head chef and the sous chef can propose the joint decision's retraction, but for reasons reviewed in Chapter 2 (§1.5), only the head chef has the power to do so using propose-and-challenge.

---

[18] You might be thinking that, unlike promises, commands are not open to challenge by their addressees at all. I address this worry in §4.3 below.

150

Finally, following the pattern of our other speech acts, we should expect commands to be invalidated by coercion and deception. I think that our intuitive judgments bear this prediction out – though it takes a bit more elaboration to see.

Begin with coercion. The claim that coercion invalidates commands is initially surprising, since many paradigm commands are backed by threats of what will happen if they are not obeyed. If the sous chef does not obey the head chef, he can be fired; if Johnson doesn't obey the sergeant, he could be court martialed. But recall from earlier that *legitimate* threats do not invalidate promises, since they can be made in good faith (Ch. 2, §2.2). Only *illegitimate* threats – i.e. those that it would be wrong to carry out – have an invalidating effect. Since the head chef can legitimately fire the sous chef for disobedience, and Johnson can be legitimately court martialed for insubordination, the deliberative theory does not predict that these threats will invalidate the commands they back.

Instead, the deliberative theory predicts that commands will be invalid when they are backed by illegitimate threats. And this seems intuitively correct:

VIOLENT CHEF: The head chef says to her sous chef, "Chop the onions! I'll punch you in the face if you don't!"

BLACKMAILING SERGEANT: The sergeant steals embarrassing photos from Johnson's personal computer. She then commands Johnson to do her tedious paperwork for her, threatening to post Johnson's embarrassing photos online if he doesn't.

These commands lack any semblance of authority. If the sous chef or Johnson were to disobey, this would not be *wrong*, even if it would be imprudent. And notably, these commands could have been authoritative if they hadn't been backed by illegitimate threats. The head chef clearly has the authority to order the sous chef to chop onions, and we can certainly imagine a case in which the sergeant could order one of her soldiers to do her paperwork for her. By making an illegitimate threat, our speakers deprive themselves of the authority they would otherwise have.

Moving to deception: intuitively, authorities do not have to reveal all of the considerations on the basis of which they make their commands. A person in authority can sometimes

legitimately keep secrets from her subjects, even when those secrets are essential to the rationale behind her commands. What the deliberative theory requires, however, is that the authority make her commands in good faith, commanding only actions that she believes to be warranted by the reasons she shares with her subjects. Thus we can predict that deception will invalidate a command when it conceals a truth that shows the command to be unwarranted, thereby demonstrating bad faith. This seems intuitively correct; consider

> **JEALOUS CHEF:** The head chef feels threatened by the talents of her sous chef. So, she hatches a plot to get the sous chef fired: she commands him to garnish the soup of a famous critic with cilantro. The head chef knows, while her sous chef does not, that the critic loathes cilantro; since the soup will be blamed on her sous chef, he will bear the brunt of the critic's fury.

> **CORRUPT SERGEANT:** The sergeant orders an attack on an enemy outpost. In fact, the sergeant knows that the attack serves no strategic purpose, and involves substantial risk to her soldiers' lives. She commands the attack only because she plans to run for political office when she returns home, and knows that the attack will make her appear tough and decisive to voters.

In my view, neither of these commands seems authoritative. If the sous chef were to discover the head chef's deception, he should feel no compunction about disobeying her. And if the soldiers discover the sergeant's true motives, they should refuse to obey her command to attack the outpost. These commands are not merely overridden by other considerations; instead, the speaker's duplicity robs them of any normative force at all.

Thus commands too are subject to the requirements of deliberative good faith. When a command is made in bad faith — whether because it is backed by an illegitimate threat or because the speaker is concealing information that shows it to be unwarranted — it fails to create any obligation for its subject to obey. This is why commands, like the other speech acts we have seen, are invalidated by certain forms of coercion and deception.

152

It is telling how similar the obligation to obey a command is to the obligations generated by promises, offers, agreements, and requests. The deliberative theory predicts this. These obligations' common features derive from their shared source: our reason to abide by our joint decisions. It is hard to see how else we might explain why commands share so many features with such seemingly distant speech acts.

4.2. The conditions of hierarchical authority

We begin, again, with the presumption against using propose-and-challenge for joint decisions concerning others' actions (Ch. 3, §3.3). Demands overcome this presumption by appeal to the weight of the shared reasons behind them. Commands require a different sort of justification. Commands propose enactive joint decisions: decisions to perform one out of several actions that are on a par. They cannot be justified by the strength of the reasons behind the action they propose, for no particular enactive decision is required by one's reasons.

For this reason, the asymmetrical power dynamic of propose-and-challenge is far more salient for commands than it is for demands. When I demand that you fulfill your obligations, it is as if I am speaking on behalf of our shared reasons. What I will is beside the point; our shared reasons (which are your reasons too) already require the action I am demanding. But when I make a command, I actively impose my will on your actions. Our shared reasons don't tell you what to do, so I tell you what to do instead. The choice of propose-and-challenge over propose-and-ratify here cries out for justification: if our shared reasons don't care whether you ϕ or ψ, then why don't I leave it up to *you* to decide which to do?

Sometimes this question has a good answer.[19] Consider

---

[19] The extant literature on hierarchical authority is overwhelmingly focused on the authority of law and the state. In contrast, I will not discuss the authority of governments at all, focusing instead on cases where particular individuals have authority to command a small number of other individuals. This is simply because addressing the authority of the state would require answering questions that would take us too far afield: what the state is, what it takes for states to be legitimate, whether the law should be understood as a set of commands, and so on. I hope to extend the theory presented here to the authority of the state in future work.

EARTHQUAKE: Disaster strikes. An earthquake hits our small town, and a school building collapses before the students can evacuate. We, a group of concerned neighbors, rush to the scene. The official rescue teams are far away, so we start trying to help. There are many jobs to be done, and they need to be done quickly. At first our group works haphazardly, but it quickly becomes clear that we need to coordinate. So one member of the group, Anna, starts giving orders: "You, in the green shirt! Go to the pharmacy down the street and get medical supplies. You two, in the red and blue, search the front of the school for survivors; the rest of you search the back. Bert, call 911 and ask when the rescue team will get here..." Relieved, we all start obeying Anna's commands.

Anna occupies no official position of authority. Before she started giving orders, she was just another concerned neighbor. And the joint decisions she is proposing are enactive: though we all have decisive reason to be helping in some way, there's no significant reason why we ought to divide up the tasks one way rather than another. (Bert could just as well have gone to the pharmacy while the green shirt called 911). Yet Anna's use of propose-and-challenge in this case – her giving *orders* rather than making requests – is clearly appropriate. Why?

The situation in EARTHQUAKE has many of the general features that justify the use of propose-and-challenge. First, there is *urgency*: we need to decide what to do quickly, since lives are at stake and time is of the essence. Second, there is a need for *decisiveness*: the cost of not acting is high, so we want the barriers to decision-making to be as minimal as possible. These factors both favor using the propose-and-challenge method. Propose-and-challenge is quick, since it does not require explicit ratification for a decision to go through, and decisive, as it errs on the side of accepting proposed decisions.

Here, the asymmetric power dynamic of propose-and-challenge is a boon as well. We need to *coordinate* our actions to be fully effective. We need a coherent joint plan that gives each of us specific roles and tasks to fulfill. But if we tried to agree on this plan by unanimous vote, for example, there would be too many proposals and opinions. After spending precious time arguing about how best to divide up our actions, we would likely arrive at a patchwork, disunified plan

that takes pieces from each of our proposals. A plan formulated by any one of us alone would likely have been better, simply because it would provide a coherent vision. So we ought to give a single person (Anna, in this case) the power to make a plan on her own and propose that plan in a way that makes it difficult to block. By putting the deliberative power in a single person's hands, the propose-and-challenge method facilitates efficient and coordinated joint action.

We can only reap these benefits if Anna *alone* is allowed to propose joint plans using propose-and-challenge. Imagine that Bert starts barking orders that are incompatible with Anna's. Confusion would ensue – whose plan are we supposed to follow? If our joint plan is to reflect a single person's coherent vision, then we need to give a single person the final say over what that plan is. This is why, once it becomes common ground that Anna is the person giving orders, it would be inappropriate for anyone else to start giving orders as well.

Thus the situation in EARTHQUAKE justifies our allowing Anna, and only Anna, to propose enactive decisions regarding our actions by means of the propose-and-challenge method. On the deliberative theory, to give Anna this deliberative power *just is* to give her the authority to command. So, the deliberative theory predicts that the situation in EARTHQUAKE is exactly the sort of situation that justifies giving someone hierarchical authority: a situation in which we need to coordinate our actions quickly and decisively under a single coherent plan.

Most hierarchical authority is not *ad hoc* in the way that Anna's authority is. Instead, most people with hierarchical authority occupy institutional roles that grant them this authority. How do we extend the deliberative theory to these institutional contexts?

The first step is to notice that the benefits of propose-and-challenge can apply to longer-term projects as well. Suppose we are building a skyscraper. Here we also have reason to follow a single person's coherent vision rather than designing by committee. (After all, we need the skyscraper to *literally* hang together). Thus we should allow a single architect to propose the building's design using the propose-and-challenge method. For the skyscraper to really represent the architect's vision, however, she will need to be able to use the propose-and-challenge method for the whole duration of the building process. But there is no prior guarantee that the architect

will have sufficient justification to use propose-and-challenge in every case. Perhaps sometimes we will think that we need to deliberate more carefully, and so will opt for propose-and-ratify. Perhaps one of us will disagree with the architect at some point and will start proposing alternative designs using propose-and-challenge.

If these possibilities are on the table, the architect's control over the skyscraper will be too fragile. When she is drafting her initial plans, for example, she will not be able to defer certain decisions of detail to later – since for all she knows, she may not be given the authority to make the decision later. If our architect is required to justify her use of propose-and-challenge on a case-by-case basis, she will not be able to fully execute her plan. To ensure that our skyscraper is based on a coherent design, then, we should agree beforehand that the architect will have the exclusive ability to employ propose-and-challenge for decisions regarding the building design for the duration of the project. This agreement gives the architect the assurance of long-term control she needs to plan and execute her vision.

Often, then, we will best harness the benefits of propose-and-challenge by making an explicit agreement to the effect that one person has the exclusive ability to use the propose-and-challenge method to propose enactive joint decisions within a delimited domain. When we make such an agreement, we thereby create an institutional role, such as the role of architect, and endow it with hierarchical authority. The person occupying this role can then formulate and execute the kind of ongoing plans that are only possible because she knows that she will continue to be in charge. By obeying her commands, we can coordinate our actions over the long term and accomplish things that would not be possible under a different, less hierarchical decision-making procedure.

Thus, I claim, a person can come to have hierarchical authority in one of two ways. First, as in EARTHQUAKE, a situation may provide us with strong reason to allow one (and only one) person to propose enactive joint plans using propose-and-challenge. Such reasons include a need for quick and decisive planning and a need for coordination. In such cases, whoever speaks up first comes to have authority, since it matters less who gives the orders than that orders are given.

In this first kind of case, the speaker's authority isn't based in any explicit contract or agreement, but instead is justified directly by the nature of the situation.

Second, as in the skyscraper example, we may explicitly agree beforehand to allow one (and only one) person to propose joint plans using propose-and-challenge within a certain domain. Such an agreement gives the relevant person the authority to command regardless of whether the reasons supporting the use of propose-and-challenge (urgency, decisiveness, coordination, etc.) apply in a particular case. I submit that the authority held by people who occupy institutional roles such as architect, head chef, sergeant, or conductor is based in agreements of this kind. These agreements are usually written into the rules or laws governing the institution: the contract by which the architect or chef is hired, the military code, the institutional norms of a university or orchestra, and so on.

These predictions of the deliberative theory match our intuitive judgments and everyday practice. Rarely is a person given the authority to issue commands without occupying a role within some explicit hierarchy, but when they are, it seems to be in exceptional situations like EARTHQUAKE where there is a clear and present need for someone to take control and coordinate action. Institutional hierarchies are also commonly justified in the way I have suggested. We have head chefs because the restaurant needs a coherent menu; we have sergeants and generals because armies need efficient and decisive direction; we have conductors because orchestras would sound awful if everyone played at the tempo of their choice. These reasons justify making a standing agreement to give a single person the right to propose enactive joint decisions using the propose-and-challenge method.

4.3. Does propose-and-challenge give enough authority?

The first objection to the deliberative theory of commands that occurs to most people is that it is not hierarchical enough. When the sergeant orders Johnson to drop and give her fifty in PUSH-UPS, she does not seem to be proposing a joint decision. If it were a *joint* decision, Johnson would have some say in it. But intuitively, Johnson has no say in the matter. Once the sergeant

157

gives the command, Johnson has to obey whether he likes it or not. Isn't this flatly at odds with the deliberative theory? No – as I now argue.

My first point is methodological. The activity of joint practical deliberation described in Chapter 1 includes, but is not limited to, the activities we call 'joint decision-making' in ordinary English. When we pay attention to the basic structure of joint practical deliberation, we come to see that joint decision-making is a more flexible and pervasive activity than we normally recognize. Thus I have argued that we should expand our conception of joint decision-making to include many things we do not ordinarily describe as such, such as joint decisions made by the propose-and-challenge method. Nothing in the structure of joint practical deliberation precludes our extending the theory in these ways, and when we do so, we find that we can illuminate phenomena that were otherwise puzzling. These theoretical fruits give us good reason to think we were right to adopt a more expansive conception of joint deliberation.

What this means is that our intuitive judgments about what to call a 'joint decision', or what interactions are naturally described as 'joint decision-making,' are not a reliable guide to what falls under the scope of joint practical deliberation as I have argued we should understand it. These direct intuitions of jointness are not a fair test of the deliberative theory. So, we should not dismiss the deliberative theory of commands just because it sounds odd to describe commands as proposing joint decisions. What matters is not how we describe commands, but whether the deliberative theory adequately captures their properties. The way to test the deliberative theory, instead, is to compare the predictions it makes about commands with the normative and communicative features we take this speech act to have.

So the question we should be asking is: does someone with the standing to propose enactive joint decisions regarding others' actions by the propose-and-challenge method have the same kind of authority as someone with the standing to issue commands? Or does the authority to command have features that are not shared by the authority to use propose-and-challenge?

With this methodology in mind, let's compare the authority given by propose-and-challenge with the authority to command. The first notable feature of propose-and-challenge is

its initial default: if a propose-and-challenge proposal is made and then nothing else is said, the proposed joint decision comes into force automatically (Ch. 1, §3.1). This feature of propose-and-challenge is clearly shared by commands. After the sergeant commands Johnson to do fifty push-ups, she need not wait for his explicit acceptance: so long as he does not object, the command immediately comes into force.

But the deliberative theory does predict that Johnson *could* object. Propose-and-challenge proposals can be challenged. Here there might seem to be a disanalogy with commands. If Johnson were to object to the sergeant's command, saying, "But Sarge, I'm eating!" this would be way out of line. It might even get Johnson punished for insubordination. Commands thus do not seem to be open to challenge in the way that the deliberative theory predicts them to be.

However, there are some situations where a command *can* be legitimately challenged. Suppose that Johnson and the sergeant are outdoors. Just as the sergeant barks, "Drop and give me fifty!" Johnson spots the enemy's planes on the horizon. Johnson could then legitimately say: "But Sarge, look: enemy planes incoming!" *This* challenge seems perfectly legitimate. Far from being insubordinate, Johnson is fulfilling his duties as a soldier by alerting the sergeant to an attack. So, it does not appear that commands are closed to challenges completely, but rather that challenges to them must meet a fairly high bar.

The deliberative theory can explain why this is so. The very considerations that justify giving someone the authority to use propose-and-challenge over others also count against raising challenges to their proposals unless absolutely necessary. As I argued in the previous section, it makes sense to give someone the authority to use propose-and-challenge when we need our joint decisions to be quick, decisive (i.e. favoring action over omission), coherent, and coordinated. These same considerations give us reason to avoid challenging the proposals of the person in authority: raising a challenge will slow us down, forestalling action and complicating the plan that is meant to coordinate our actions. In an institution such as the military where quick, decisive, and coordinated action is essential, we should expect there to be strong norms discouraging challenges to the decisions of those in authority. To justify challenging a command,

159

then, one must have something important to say – such as that there are enemy planes on the horizon. To challenge a command with insufficient justification is to disregard the importance of the considerations that justify the commander's authority in the first place. This is why Johnson could be punished for raising a trivial challenge like "But Sarge, I'm eating!"

The truth, then, is not that commands cannot be challenged, but that to challenge a command, one needs good reason to do so. Not just any objection will do. And this is just what the deliberative theory would lead us to expect, given the kind of situations in which we grant people hierarchical authority.

The final relevant feature of propose-and-challenge what we earlier called its *final default* (Ch. 1, §3.1). In propose-and-challenge, if there is disagreement between the speaker and the addressee(s) over whether to accept a proposed joint decision, this disagreement is resolved by default in favor of the proposal, bringing it into force. This is the most important way in which propose-and-challenge gives the speaker deliberative power. It means that, if the speaker is unconvinced in good faith by an addressee's challenges, then her proposed joint decision will go through whether the addressee likes it or not. In short, the speaker has the last word over whether the proposal is accepted: if she stands by her proposal, it will come into effect.

This fits commands nicely. Whatever objections Johnson may raise, the sergeant can respond, "That's final" or, "That's an order." And if she says this, then there really is nothing Johnson can do to prevent her command from coming into force. This is the truth behind the thought that Johnson has 'no say' regarding the sergeant's command. It's not that he has no say – he can challenge the command – but he does not have the *final* say. Ultimately, it is up to the sergeant to decide whether the joint decision her command proposes will come into force.

While giving the speaker the last word in this way matches our everyday practice with commands, you might worry that it undermines the idea that propose-and-challenge is aptly described as making *joint* decisions. If the sergeant can bulldoze over any of Johnson's objections, then can we really say that the result of her command is a decision she made jointly *with Johnson?*

Yes, for two reasons. First, Johnson still has a role to play, even if he has less deliberative power than the sergeant. He is given an opportunity to evaluate the command and raise any challenges he thinks sufficiently serious to be worth bringing up. Second, and more importantly, the sergeant's commands must be based upon, and constrained by, the reasons she shares with Johnson. In other words, the sergeant must hold her commands accountable to the norms that apply to joint decisions. To deliberate in good faith, the sergeant must make only commands that she believes to be warranted by the reasons she and Johnson share. And if Johnson does challenge her command, the sergeant has to take his challenge seriously. If she judges that Johnson's challenge shows the joint decision proposed by her command to be unwarranted, she must retract it. These requirements of good faith distinguish the authority to command from the brute imposition of one's will by force: the authority to command is a kind of power over the process of joint deliberation, and so is subject to the norms of this activity.

On reflection, we seem to hold authorities accountable to the requirements of good faith. People in positions of authority are expected to exercise their authority in the service of the purposes for which the authority was given – i.e., the reason-giving projects they share with their subordinates. When authorities violate this norm, exercising their powers to serve their own personal whims or interests, we take this to undermine their claim to authority. We have already seen this in the cases of coercive and deceptive commands we considered in §4.1: the fact that these commands were made in bad faith robs them of their normative force. More generally, when we find out that a person in authority is abusing their office for personal gain, neglecting the duties that they were put in power to perform, our response is to cease respecting that person's claim to authority altogether. We will not hesitate to disobey their commands.

Thus the deliberative theory helps us to explain why we take corruption to undermine a person's claim to authority: because it shows they are not exercising that authority in good faith. Contrast an authority who makes commands in good faith, but has poor judgment. When an authority makes misguided commands, we might think that she should be taken out of the position of authority, but so long as she remains in this position, we still feel bound to obey her

commands. Corruption invalidates authority in a way that mere mistakes of judgment do not —
just as the deliberative theory would lead us to expect.

A person with the standing to propose joint decisions regarding other's actions by means
of propose-and-challenge will have the following kind of authority. Her proposals will bind the
addressee's actions by default, unless the addressee raises a legitimate challenge to them. Due to
the kinds of situations in which we have reason to give someone authority, however, there will
typically be a strong presumption against challenging her proposals unless necessary. And when
the addressee *does* challenge a proposal, the speaker still has the last word: if she judges in good
faith that the challenge is not decisive, she can reject it and bring her proposed joint decision into
force. The speaker's exercise of this power over their joint deliberation is constrained only by the
norms of good faith, which require her to make only proposals that she believes to be warranted
by the shared reasons, and retract her proposal if a challenge shows it to be unwarranted.

I submit that this is the same kind of authority that comes with the power to command.
Some people have certainly claimed more authority than this for themselves, asserting that their
commands may never be challenged, or that they may exercise their authority arbitrarily,
commanding whatever they wish for any reason they please. But I see no reason why we should
recognize these claims as legitimate. Though it is useful to have contexts in which challenges to
an authority are rare, to make an authority's commands unchallengeable *no matter what*, even
when the challenge is important, would give us little gain in coordination or efficiency at the
steep cost of sometimes missing out on critical information ("enemy planes incoming!"). And
while we may have reason to give some people broad domains of authority within which they can
exercise wide discretion (e.g., a parent's authority over her children), to release those people
entirely from the requirements of good faith would be to erase the distinction between legitimate
authority and brute domination. I conclude that the deliberative theory gives persons in positions
of hierarchical authority as much authority as we reflectively judge they ought to have.

4.4. Comparison with Raz's theory

Joseph Raz's theory of hierarchical authority (summarized in §1.2.1 of this chapter) has generated an enormous amount of commentary (for a review, see Ehrenberg 2011). Much of this literature has been critical, raising objections and counterexamples that, I believe, show the theory to be untenable as it is formulated. In spite of these criticisms, however, Raz's theory remains in the foreground of the debate because it seems to many to contain deep insights about the nature of hierarchical authority. Here I will argue that the deliberative theory captures the truth in Raz's theory while avoiding its problems.

Recall that Raz's theory is composed of three major theses: the normal justification thesis, the dependence thesis, and the preemption thesis. The deliberative theory vindicates partial versions of each of these theses.

Begin with the preemption thesis, which states: "The fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of some of them" (Raz 1986: 46). I think this thesis is partly false and partly true. It is false if we interpret it as describing the subject's actual normative reasons. When the sergeant commands Johnson to do fifty push-ups, his other normative reasons to do push-ups (e.g., that push-ups make him stronger) do not disappear. More importantly, the reason provided by the sergeant's command does not merely replace Johnson's prior reasons to do fifty push-ups, leaving the total balance of reasons unchanged. If it did, then Johnson would have no more reason to do push-ups after the command is made than he had before. Clearly, the command creates a new reason for Johnson to do push-ups that *is* "added to all other relevant reasons" (46).

The truth in the preemption thesis is that the addressee of a command should not *think* of the command as offering just another reason in favor of doing the action commanded. This is because the addressee of a legitimate command should not be weighing reasons for and against obeying the command at all. Instead, if the command is authoritative, the addressee should treat it as already decided that he will do the action commanded, and thus should not need to weigh

the reasons for and against doing it. As we have seen, the deliberative theory explains this preemptive effect of commands as resulting from the deliberation-closing force of joint decisions.

The dependence thesis says: "All authoritative directives should be based on reasons which already independently apply to the subjects of the directives and are relevant to their action in the circumstances covered by the directive" (47). The deliberative theory vindicates this thesis without qualification. Commands, like any other proposals in joint deliberation, must be based upon the shared reasons of the speaker and addressee to be made in good faith. And, applying the UNANIMITY principle defended in Ch. 1 (§1.3 and §5.1), we can infer that any shared reason of the speaker and addressee must also be a reason within the addressee's individual practical deliberation. So the deliberative theory entails that good faith commands must be based upon reasons that apply independently to their addressees.[20]

The heart of Raz's theory, and the target of the most criticism, is the normal justification thesis. The normal justification thesis claims that the "normal way to establish that a person has authority over another person" is to show that the subject of that authority would better comply with his own reasons for action by following the authority's commands than he would by deciding what to do on his own (53). The deliberative theory mostly agrees: a large part of the justification for giving you the authority to propose enactive joint decisions regarding my actions by means of propose-and-challenge is that this will make me better able to act in accordance with my reasons – and in particular, the reasons I share with you. As we discussed in §4.2, different factors might make this the case: perhaps we simply need to decide quickly, perhaps we need to coordinate our actions, perhaps our actions need to be guided by a coherent vision, perhaps you are just a better judge of the reasons in this domain than I am. Any of these factors can contribute to the justification for giving someone hierarchical authority.

We can see where the deliberative theory departs from the normal justification thesis by considering the objections that have been raised to it. First, some have complained that the

---

[20] As I pointed out in the previous section, this explains why commands lack authority when they are based on the personal whims of the person in authority rather than on the rightful priorities of her office.

normal justification thesis makes authority too *fragile* (e.g., Durning 2003: 603-604). It seems to predict that, when a subject of authority would better comply with her reasons by judging what to do for herself than by obeying the authority's commands, the commands have no force. But this does not fit our practice. A sous chef is subject to the head chef's authority even if, for a particular dish, he would make it better by following his own recipe than by following the head chef's orders. The head chef does not lose her authority just when the sous chef happens to disagree with her – even if the sous chef is right. The reason why authority is not fragile in this way is because of the role of agreements in creating positions of authority. If we have *agreed* that the head chef is in charge of the kitchen, then this agreement obliges the sous chefs to obey her commands even when they would make better food on their own. Authority-granting agreements thus enable a person to have hierarchical authority even when, in a particular case, the normal justification criterion is not met. As we saw in the architect example, the benefits of giving one person assured control within a domain can outweigh the cost of sometimes having to obey a misguided command.

The second major complaint about the normal justification thesis is that it seems to *overgenerate*, granting authority in too many cases (Hershovitz 2011; Darwall 2013a). There are many cases in which I would better comply with my reasons by obeying your direction than by deciding myself, and yet you intuitively have no authority to command me. Consider

> **COOKING:** Cora is trying to cook a traditional Chinese dish. She wants to make it as well as she can. As it happens, her roommate John is an expert on Chinese cuisine, while Cora knows little about it. Unbidden, John walks into the kitchen and starts issuing commands: "Turn up the heat! Add more red pepper!"[21]

Darwall argues, and I agree, that John has no authority to tell Cora what to do in this case. And yet his commands clearly meet the criteria of the normal justification thesis: since Cora wants to prepare the dish as well as possible, and John knows far better than Cora how to do this, she will

---

[21] Though Darwall brought it to my attention (2013a: 147), this example is actually Raz's own (1986: 64).

better comply with her reasons for action by following John's commands than by deciding what to do herself. But this does not seem sufficient to give John authority over Cora.

For similar reasons, the normal justification thesis also fails to accommodate the domain-restrictedness of hierarchical authority. As I have said, the head chef can command her sous chefs in the restaurant kitchen, but she has no authority over them when they are cooking at home. Similarly, a professor can order her students to read *The Republic* while they are in her class, but after the semester ends, she has no say over what her students read. Authority can also be given or taken away by explicit acts. If the head chef is fired, then she immediately loses her authority to command the sous chefs. When your fellow soldier is promoted to a rank above you, she suddenly has an authority over you that she did not have before.

If we think, as the normal justification thesis suggests, that the existence of authority depends primarily on a person's being better placed to judge the relevant reasons for action, then these observations should be surprising. The head chef may be just as good at judging how her sous chefs ought to cook at home as she is at judging how they ought to cook in the restaurant – so why does she only have authority in the latter case? When your friend is promoted to sergeant, the mere fact that she has been given a new rank does not make her any better at judging what you ought to do – so why does she now have authority over you?

Again, the deliberative theory explains these judgments by appeal to the mediating role of explicit agreements in grounding institutional positions of authority. The head chef's authority over her sous chefs is based in an agreement to the effect that she can command the sous chefs while they are in the restaurant. This agreement specifies a particular domain for the head chef's authority that, thankfully, does not extend to the sous chefs' home kitchens. And when the head chef is fired, the agreement is thereby retracted, and thus no longer grants her the authority it once did. Generally, the deliberative theory predicts that role-based authority (a) will be restricted to the domain covered by the authority-granting agreement, and (b) can be explicitly granted or revoked by means of making or retracting this agreement. These predictions seem to fit the facts.

What about COOKING? This example illustrates just how much it takes to justify enactive uses of propose-and-challenge over others' actions. Yes, Cora will cook better if she obeys John's commands. But Cora also has a strong interest in having control over her own actions. The prospect of higher-quality cooking is not enough to outweigh the cost to Cora of accepting the unbalanced power dynamic involved in giving authority to John. Plus, while John knows more than Cora about what to do, the other factors that justify uses of propose-and-challenge are not present. There is not the urgency and need for decisiveness that justifies Anna's authority in EARTHQUAKE; there is not the need for a unified vision that justifies the architect's authority over the skyscraper. Thus there is little reason why John should assist Cora by making *commands* rather than, say, giving her advice or offering his help. If the situation were different – if, say, instead of cooking Chinese, Cora was defusing a bomb that was about to detonate – then John might come to have the authority to give her orders. (This would be the *ad hoc* kind of authority we saw in EARTHQUAKE). But the fact that John knows better than Cora what she should do is not sufficient on its own to give him authority over her.

I conclude that the deliberative theory better fits our judgments about the extension of hierarchical authority than Raz's account, while retaining the motivating insights that have rightfully earned Raz's theory a central place in the literature.

5. The redundancy argument redux

At the end of Chapter 2, I raised a challenge for alternatives to the deliberative theory of promises. Our theory of joint practical deliberation predicts that a certain kind of proposal in this activity will have all the characteristic features of promises. Thus the opponent of the deliberative theory faces a dilemma: she must either accept the deliberative theory's claim that promises are identical to these deliberative proposals, or accept that promises are redundant. The latter option is deeply unattractive, I argued; so they should opt for the former. We are now in a position to apply the same argument to demands and commands.

One of the categories of deliberative proposal generated by our theory of joint practical deliberation is a proposal to make a joint decision regarding the addressee's actions by means of the propose-and-challenge method. We can make a further distinction within this category between proposals to make commissive joint decisions and proposals to make enactive joint decisions. When we take a careful look at these speech acts' properties, I have argued, we will find that they are identical to the properties of demands and commands, respectively.

These two kinds of deliberative proposal share the features that demands and commands have in common. By proposing a joint decision, both purport to close the addressee's deliberation in a way that advice and threats do not (§1.1). Because they employ the propose-and-challenge method, they do not require the addressee's explicit acceptance to come into force, instead requiring only recognition and the absence of successful challenges.

Distinguishing between commissive and enactive proposals generates two speech act categories that differ in the same ways as demands and commands. Commissive proposals put forward joint decisions that the addressee is already required by the shared reasons to perform. Thus, like demands, these proposals hold their addressees accountable to reasons they already have. Moreover, by looking at what it takes for our shared reasons to require an action, we can predict when others will have the standing to propose commissive joint decisions about our actions by means of propose-and-challenge (§3.1). When we compare these predictions to our judgments about when others have standing to make demands, the two appear to match (§3.2). No one has standing to propose in this way that I floss, because my reasons to floss are not shared to any significant extent with others; everyone has standing to propose in this way that I not dump poison in the ocean, because my reasons not to do so are shared by all persons.

Thus, by proposing a commissive joint decision regarding someone else's actions by means of the propose-and-challenge method, we can hold them accountable in the same way, under the same conditions, as we do by making a demand. The best explanation of this observation seems to be that demands are identical to this commissive kind of deliberative

proposal. If the two are distinct, then demands are redundant: everything we do with demands, we could do with proposals in joint practical deliberation instead.

Our second category is a proposal to make an *enactive* joint decision regarding the addressee's actions by means of the propose-and-challenge method. Enactive joint decisions decide between shared options that were previously on a par, choosing one to be performed at the exclusion of the others. By bringing an enactive joint decision into force, this second kind of proposal creates a new obligation that has the same normative upshots as the obligation created by a command (§4.1). Since these proposals give the speaker an asymmetrical power to impose her will on the addressees' actions, they need strong justification to be used appropriately. The conditions under which such proposals are appropriate are tellingly similar to the conditions under which commands are appropriate (§4.2). In situations like EARTHQUAKE, where a group of people needs to coordinate their actions in a quick, efficient, and decisive manner, we have reason to give one person the ability to propose enactive joint decisions using propose-and-challenge – in just the same way as we have reason to give someone the authority to make commands. In other situations, as when we are building a skyscraper, running a restaurant, or forming an army, we might want to explicitly agree to give someone this asymmetrical deliberative power on a more permanent basis, so they can shape our joint undertaking under a coherent plan. These agreements granting someone a right to make enactive propose-and-challenge proposals within a certain domain are similar in both their rationale and effect to the agreements by which we create institutional roles of hierarchical authority.

Thus, by proposing an enactive joint decision regarding someone else's actions by means of the propose-and-challenge method, we can put them under a new obligation in the same way, under the same conditions, as we do by giving a command. The best explanation of this observation seems to be that commands are identical to this enactive kind of deliberative proposal. If the two are distinct, then commands are redundant: everything we do with commands, we could do with proposals in joint practical deliberation instead.

If our opponent decides to bite the bullet and accept that commands and demands *are* redundant, she faces the same awkward questions we noted in Chapter 2. If commands and demands are redundant, then why do we have them at all? Why do we use commands to coordinate action, when proposing enactive joint decisions using propose-and-challenge would do the job just as well? Why do we use demands to hold one another accountable, when proposing commissive joint decisions by propose-and-challenge would have the same effect? How have these two categories of deliberative proposal, both with significant social and moral effects, escaped the notice of both folk observation and philosophical inquiry? And finally, if commands and demands are distinct from these two kinds of deliberative proposal, then why do they have the same properties? Is it yet another remarkable coincidence?

When we accept the deliberative theory, these perplexing mysteries disappear. Demands and commands are so similar to proposals in joint practical deliberation because they *are* proposals in joint practical deliberation. There is no coincidence or similarity to be explained: we have simply discovered an identity.

Once we accept this identity, we gain the full explanatory resources of the deliberative theory. We can now *explain* why others cannot demand that I floss by noting that our shared reasons do not give me decisive reason to do so. We can explain why Anna has the authority to command in EARTHQUAKE, while John has no such authority in COOKING, by appeal to the costs and benefits of the propose-and-challenge method. And so on. As we draw our map of joint practical deliberation, we discover a natural place for demands and commands. When we locate demands and commands on this map, they begin to make more sense: we can better see why we make these speech acts, why they have the effects we take them to have, and what it takes to employ them legitimately. This gives us good reason to think that the map is right.

Conclusion

Three Cheers for the Deliberative Theory

The deliberative theory of promises, offers, agreements, requests, demands, and commands claims that each of these speech acts is a move in joint practical deliberation, the activity of deciding together what to do. These speech acts map on to the different possible moves in joint deliberation as represented in the following table:

	Propose-and-challenge	**Propose-and-ratify**
"I will φ"	Promises	Offers
"You will φ"	Commands & Demands	Requests
"We will φ"	Agreements	

In this dissertation, I have offered three general reasons for thinking that these analyses are true.

The first reason I called the *redundancy argument* (Ch. 2, §5; Ch. 4, §5). When we look at the speech acts we would need to engage in joint practical deliberation, we find that they match up exactly with various speech acts we recognize in ordinary language. Applying our theory of joint deliberation, we can predict that these deliberative proposals will have properties that exactly match those of their ordinary language twins. A proposal to make a joint decision regarding one's own actions by means of the propose-and-challenge method has the same normative effects and validity conditions as a promise. The properties of proposals to make joint decisions regarding an addressee's actions by means of propose-and-ratify exactly align with the intuitive properties of requests. A proposal to make an enactive joint decision concerning the addressee's actions by the propose-and-challenge method is indistinguishable from a command. And so on.

171

If my arguments for these claims have been sound, then it is hard to justify denying that these deliberative proposals are identical to the ordinary speech acts with which they correspond. Doing so would commit you to thinking that all of these ordinary speech acts are redundant. Everything we accomplish using demands, offers, promises, and so on, we could do instead by making proposals in joint practical deliberation. Thus the opponent of the deliberative theory faces some stubborn puzzles. Why do we have these speech acts, if proposals in joint deliberation would work just as well? Why are these speech acts all so similar to acts within joint deliberation? Why are requests, for example, so similar to proposals to make joint decisions about others' actions by the propose-and-ratify method? And why do we not have ordinary language words for any of the deliberative proposals in this set? How could all these admissible ways of deciding together have escaped our notice entirely? These problems have a clear solution: accept that our ordinary speech acts are identical to their corresponding proposals in joint deliberation. There are not two sets of speech acts here, but one.

The second major reason for accepting the deliberative theory I called the *unity argument* (Ch. 3, §4). This argument points to the similarities between our ordinary speech acts as also in need of explanation. Promises, offers, agreements, requests, demands and commands all bring obligations into force that share a telling number of features. Each of these acts is invalidated by coercion and deception between the speaker and addressee. All of these acts generate directed obligations that hold between the speaker and addressee in particular. These acts all have exclusionary effects on deliberation, closing the question of whether a certain action will be done and making it inappropriate for the agent to reopen that question in later deliberation. These shared patterns between apparently disparate phenomena call out for explanation.

The deliberative theory explains our speech acts' common features readily. These speech acts all have the same primary effect: to bring a joint decision into force between the speaker and her addressee(s). We can thus explain their similar normative upshots by appeal to the normative upshots of making a joint decision, and their similar validity conditions by appeal to the norms of deliberative good faith.

172

Our opponent has a more difficult time explaining the similarities between these speech acts. Again and again, we have seen efforts to explain these similarities by reducing one speech act to another fail. Offers, agreements, and requests cannot be adequately understood as special cases of promises. Commands cannot be reduced to demands; nor can demands be reduced to commands. The failure of these reductive views leaves the opponent of the deliberative theory with few resources with which to explain the unity of our target speech acts. Again, the solution seems clear: these speech acts are unified by the fact that they are all moves in joint deliberation.

While I find these two arguments compelling, I must admit that I was already convinced of the deliberative theory before I formulated either the redundancy or unity arguments. What sold me on the deliberative theory was not some devastating objection to its opponents. Instead, I became convinced of the deliberative theory when I saw how, again and again, it helped me come to understand otherwise puzzling features of its analysanda.

This is the third and most important reason why the deliberative theory deserves our acceptance: its explanatory power. From the minimal constraints on which we based our account of joint practical deliberation, we can derive a theory that captures and explains the shape and significance of our ordinary practices of promising, offering, agreeing, requesting, demanding and commanding. We can explain why illegitimate threats nullify promises, while legitimate threats do not, by thinking about what is required to deliberate in good faith. We can capture the painful awkwardness of turning down a request without a good excuse by thinking about what rejecting a deliberative proposal implies about your beliefs. We can explain why my partner can demand things of me that my neighbor cannot by reflecting on what reasons we share within our different deliberative perspectives. We can understand when and why some people are given the authority to command others by thinking about the costs and benefits of using the propose-and-challenge method. These and the many other explanations the deliberative theory has generated along the way constitute the strongest case for its truth.

Once we start thinking of these speech acts as proposals in joint practical deliberation, they no longer seem like mysterious acts of changing the moral law by declaration. They become

predictable parts of our daily activity of figuring out how to live together. Under the light of the deliberative theory, promises, offers, agreements, requests, demands, and commands start to make sense. We can locate them in the world we live in. They lose their magic, becoming expected, obvious, even mundane.

Isn't that what theories are supposed to do?

References

Anscombe, G. E. M. 1978. Rules, Rights, and Promises. *Midwest Studies in Philosophy* 3:318-323.

Árdal, Páll S. 1968. "And That's a Promise." *Philosophical Quarterly* 18:225-237.

Austin, J. L. 1962. *How To Do Things With Words*. Oxford: Oxford University Press.

Baier, Kurt. 1966. Moral Obligation. *American Philosophical Quarterly* 3: 210-226.

Block, Ned, and Robert Stalnaker. 1999. Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review* 108:1-46.

Brandom, Robert. 1983. Asserting. *Noûs* 17:637-650.

Bratman, Michael E. 1987. *Intentions, Plans, and Practical Reason*. Stanford: CSLI Publications.

Bratman, Michael E. 2007. Three Theories of Self-Governance. In *Structures of Agency*, 222-253. Oxford: Oxford University Press.

Bratman, Michael E. 2009. Intention, Practical Rationality, and Self-Governance. *Ethics* 119:411-443.

Bratman, Michael E. 2014. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press.

Cappelen, Herman. 2011. Against Assertion. In *Assertion: New Philosophical Essays*, edited by Jessica Brown and Herman Cappelen, 21-46. Oxford: Oxford University Press.

Chang, Ruth. 2002. The Possibility of Parity. *Ethics* 112:659-688.

Chwang, Eric. 2011. On Coerced Promises. In *Promises and Agreements*, edited by Hanoch Sheinman, 156-182. Oxford: Oxford University Press.

Corbin, Arthur. 1917. Offer and Acceptance, and Some of the Resulting Legal Relations. *Yale Law Journal* 26:169-206.

Craswell, Richard. 1996. Offer, Acceptance, and Efficient Reliance. *Stanford Law Review* 48:481-553.

Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge: Harvard University Press.

Darwall, Stephen. 2011. Demystifying Promises. In *Promises and Agreements*, edited by Hanoch Sheinman, 255-276. Oxford: Oxford University Press.

Darwall, Stephen. 2012. Bipolar Obligation. In *Oxford Studies in Metaethics, Vol. 7*, edited by Russ Shafer-Landau, 333-356. Oxford: Oxford University Press.

Darwall, Stephen. 2013a. Authority and Second-Personal Reasons for Acting. In *Morality, Authority and Law: Essays in Second-Personal Ethics I* (pp. 135-150). Oxford: Oxford University Press.

Darwall, Stephen. 2013b. Authority and Reasons: Exclusionary and Second-Personal. In *Morality, Authority and Law: Essays in Second-Personal Ethics I* (pp. 151-167). Oxford: Oxford University Press.

Deigh, John. 2002. Promises Under Fire. *Ethics* 112:483-506.

Dougherty, Tom. 2013. Sex, Lies, and Consent. *Ethics* 123:717-744.

Dummett, Michael. 1993. Language and Truth. In *The Seas of Language* (pp. 117-146). Oxford: Oxford University Press.

Durning, Patrick. 2003. Joseph Raz and the Instrumental Justification of a Duty to Obey the Law. *Law and Philosophy* 22:597-620.

Ehrenberg, Kenneth. 2011. Critical Reception of Raz's Theory of Authority. *Philosophy Compass* 2011: 777-785.

Enoch, David. 2009. Wouldn't It Be Nice If *p*, Therefore, *p* (for a moral *p*). *Utilitas* 21:222-224.

Evans, Mike. 1996. Gifts and Commodities on a Tongan Atoll: Understanding Intention and Action in a MIRAB Economy. Unpublished dissertation.
<https://macsphere.mcmaster.ca/bitstream/11375/7026/1/fulltext.pdf>

Frankfurt, Harry. 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68: 5-20.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge: Harvard University Press.

Gilbert, Margaret. 1993a. Agreements, Coercion, and Obligation. *Ethics* 103:679-706.

Gilbert, Margaret. 1993b. Is an Agreement an Exchange of Promises? *Journal of Philosophy* 90:627-649.

Gilbert, Margaret. 2004. Scanlon on Promissory Obligation: The Problem of Promisees' Rights. *Journal of Philosophy* 101:83-109.

Gilbert, Margaret. 2011a. Acting Together. In *Joint Commitment: How We Make the Social World*, 23-36. Oxford: Oxford University Press.

Gilbert, Margaret. 2011b. Three Dogmas about Promising. In *Promises and Agreements*, edited by Hanoch Sheinman, 83-108. Oxford: Oxford University Press.

Grice, Paul. 1989. Logic and Conversation. In *Studies in the Way of Words*, 22-40. Cambridge: Harvard University Press.

Hart, H. L. A. 1955. Are There Any Natural Rights? *Philosophical Review* 64: 175-191.

Hart, H. L. A. 1982. Commands and Authoritative Legal Reasons. In *Essays on Bentham: Jurisprudence and Moral Theory*, 243-268. Oxford: Oxford University Press.

Hedden, Brian. 2012. Options and the Subjective *Ought*. *Philosophical Studies* 158:343-360.

Hershovitz, Scott. 2011. The Role of Authority. *Philosophers' Imprint* 11:1-19.

Holton, Richard. 2004. Rational Resolve. *Philosophical Review* 113:507-535.

Hooker, Brad. 2011. Promises and Rule-Consequentialism. In *Promises and Agreements*, edited by Hanoch Sheinman, 237-254. Oxford: Oxford University Press.

Hume, David. 1739/1978. *A Treatise of Human Nature*, edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press.

Jones, David H. 1966. Making and Keeping Promises. *Ethics* 76:287-296.

Kamm, Frances. 1992. Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs* 21:354-389.

Kant, Immanuel. 1785/1996. *Groundwork of the Metaphysics of Morals*. In *Practical Philosophy*, translated and edited by Mary J. Gregor, 37-108. Cambridge: Cambridge University Press.

Kavka, Gregory. 1983. The Toxin Puzzle. *Analysis* 43:33-36.

de Kenessey, Brendan. What Is Obligation? Unpublished manuscript.

Kolodny, Niko, and R. Jay Wallace. 2003. Promises and Practices Revisited. *Philosophy and Public Affairs* 31:119-154.

Korn, Fred, and Shulamit R. Decktor Korn. 1983. Where People Don't Promise. *Ethics* 93:445-450.

Korsgaard, Christine M. 1986. The Right to Lie: Kant on Dealing with Evil. *Philosophy and Public Affairs* 15:325-349.

Korta, Kepa and John Perry. 2015. Pragmatics. In *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), edited by Edward N. Zalta.
<http://plato.stanford.edu/archives/sum2015/entries/pragmatics/>.

Laden, Anthony Simon. 2012. *Reasoning: A Social Picture*. Oxford: Oxford University Press.

Langton, Rae. 1992. Duty and Desolation. *Philosophy* 67:481-505.

Lewis, David. 1969/2002. *Convention: A Philosophical Study*. Oxford: Blackwell Publishers Ltd.

Lewis, David. 1979a. A Problem about Permission. In *Essays in Honour of Jaakko Hintikka*, edited by Esa Saarinen et al., 163-175. Dordrecht: D. Reidel Publishing Company.

Lewis, David. 1979b. Scorekeeping in a Language Game. *Journal of Philosophical Logic* 8:339-359.

Liberto, Hallie. Promises and the Backward Reach of Uptake. Forthcoming in *American Philosophical Quarterly*.

Locke, John. 1689/1980. *Second Treatise of Government*. Edited by C. B. Macpherson. Indianapolis: Hackett Publishing Company, Inc.

MacCormick, Neil. 1972. Voluntary Obligations and Normative Powers – I. *Proceedings of the Aristotelian Society, Supplementary Volumes* 46:59-78.

Macnamara, Coleen. 2011. Holding Others Responsible. *Philosophical Studies* 152:81-102.

Maitra, Ishani. 2011. Assertion, Norms, and Games. In *Assertion: New Philosophical Essays*, edited by Jessica Brown and Herman Cappelen, 277-295. Oxford: Oxford University Press.

McGovney, D. O. 1914. Irrevocable Offers. *Harvard Law Review* 27:644-663.

McMahon, Christopher. 1989. Promising and Coordination. *American Philosophical Quarterly* 26: 239-247.

Nagel, Thomas. 1995. Personal Rights and Public Space. *Philosophy and Public Affairs* 24:83-107.

Narveson, Jan. 1971. Promising, Expecting, and Utility. *Canadian Journal of Philosophy* 1:207-233.

O'Neill, Onora. 1985. Between Consenting Adults. *Philosophy and Public Affairs* 14:252-277.

Oliphant, Herman. 1920. The Duration and Termination of an Offer. *Michigan Law Review* 18:201-212.

Owens, David. 2006. A Simple Theory of Promising. *Philosophical Review* 115:51-77.

Owens, David. 2007. Duress, Deception, and the Validity of a Promise. *Mind* 116:293-315.

Owens, David. 2008. Promising Without Intending. *Journal of Philosophy* 105:737-755.

Owens, David. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press.

Owens, David. 2014. Does a Promise Transfer a Right? In *Philosophical Foundations of Contract Law*, edited by George Letsas, Prince Saprai and Gregory Klass (pp. 78-95). Oxford: Oxford University Press.

Pallikkathayil, Japa. 2011. The Possibility of Choice: Three Accounts of the Problem with Coercion. *Philosophers' Imprint* 11: 1-20.

Pink, Thomas. 2009. Promising and Obligation. *Philosophical Perspectives* 23: 389-420.

Preston-Roedder, Ryan. 2014. A Better World. *Philosophical Studies* 168:629-644.

Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. 2004. The Strike of the Demon: On Fitting Pro-attitudes and Value. *Ethics* 114:391-423.

Rawls, John. 1955. Two Concepts of Rules. *Philosophical Review* 64:3-32.

Rawls, John. 1971. *A Theory of Justice*. Cambridge: Belknap Press.

Raz, Joseph. 1972. Voluntary Obligations and Normative Powers – II. *Proceedings of the Aristotelian Society, Supplementary Volumes* 46:79-102.

Raz, Joseph. 1975. *Practical Reason and Norms*. Oxford: Oxford University Press.

Raz, Joseph. 1977. Promises and Obligations. In *Law, Morality, and Society: Essays in Honour of H. L. A. Hart*, edited by P. M. S. Hacker and Joseph Raz, 210-228. Oxford: Clarendon Press.

Raz, Joseph. 1985. Authority and Justification. *Philosophy & Public Affairs* 14:3-29.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.

Raz, Joseph. 2006. The Problem of Authority: Revisiting the Service Conception. *Minnesota Law Review* 90:1003-1044.

Raz, Joseph. 2010. On Respect, Authority, and Neutrality: A Response. *Ethics* 120:279-301.

Rescorla, Michael. 2009. Assertion and its Constitutive Norms. *Philosophy and Phenomenological Research* 79: 98-130.

Ripstein, Arthur. 2009. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge: Harvard University Press.

Robins, Michael H. 1984. *Promising, Intending, and Moral Autonomy*. New York: Cambridge University Press.

Roth, Abraham Sesshu. 2016. Intention, Expectation, and Promissory Obligation. *Ethics* 127:88-115.

Scanlon, T. M. 1990. Promises and Practices. *Philosophy and Public Affairs* 19:199-226.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Sheinman, Hanoch (ed.). 2011. *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press.

Shiffrin, Seana Valentine. 2008. Promising, Intimate Relationships, and Conventionalism. *Philosophical Review* 117:481-524.

Shiffrin, Seana Valentine. 2011. Immoral, Conflicting, and Redundant Promises. In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 155-178. Oxford: Oxford University Press.

Shockley, Kenneth. 2007. On that Peculiar Practice of Promising. *Philosophical Studies* 140:385-399.

Smith, Michael. 2011. The Value of Making and Keeping Promises. In *Promises and Agreements*, edited by Hanoch Sheinman, 198-216. Oxford: Oxford University Press.

van Someren Greve, Rob. 2011. Wishful Thinking in Moral Theorizing: Comment on Enoch. *Utilitas* 23:447-450.

Southwood, Nicholas, and Daniel Friedrich. 2009. Promises Beyond Assurance. *Philosophical Studies* 144: 261-280.

Stalnaker, Robert C. 1984. *Inquiry*. Cambridge: The MIT Press.

Stalnaker, Robert C. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Taylor, Erin. 2013. A New Conventionalist Theory of Promising. *Australasian Journal of Philosophy* 91:667-682.

Thomson, Judith Jarvis. 1990. *The Realm of Rights*. Cambridge: Harvard University Press.

Tuomela, Raimo. 2007. *The Philosophy of Sociality: The Shared Point of View*. Oxford: Oxford University Press.

Wallace, R. Jay. 2013. The Deontic Structure of Morality. In *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*, edited by David Bakhurst, Brad Hooker, and Margaret Olivia Little, 137-167. Oxford: Oxford University Press.

Watson, Gary. 1975. Free Agency. *The Journal of Philosophy* 72: 205-220.

Watson, Gary. 2003. The Work of the Will. In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud and Christine Tappolet (pp. 172-200). Oxford: Oxford University Press.

Westlund, Andrea C. 2009. Deciding Together. *Philosophers' Imprint* 9:1-17.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Wittgenstein, Ludwig. 1953/2009. *Philosophical Investigations* (Revised 4th Edition). Translated by G. E. M. Anscombe, edited by P. M. S. Hacker and Joachim Schulte. West Sussex: Wiley-Blackwell.

Yablo, Stephen. A Problem about Permission and Possibility. In *Epistemic Modality*, edited by Andy Egan and Brian Weatherson, 270-294. Oxford: Oxford University Press.