An Update Semantics for Promises and Other Obligation-Creating Speech Acts

A Promising Start

MSc Thesis *(Afstudeerscriptie)*

written by

Aadil Hanif Kurji
(born October 11th, 1983 in Calgary, Canada)

under the supervision of **Prof Dr Frank Veltman**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
October 9, 2012	Prof Dr Benedikt Löwe
	Prof Dr Martin Stokhof
	Prof Dr Frank Veltman
	Dr Alexandru Baltag
	Dr Robert van Rooij

Institute for Logic, Language and Computation

Abstract

This thesis is concerned with developing an update semantics to model the speech act of promising, under various philosophical frameworks.

There are two main families of promissory framework, conventional and expectational, we offer a hybrid account which makes up for the deficiencies found with both. We also discuss promises in relation to other obligation-creating speech acts to derive a set of desiderata for the formalization. Promises are speech acts which have specific illocutionary and perlocutionary effects which directly map to the deontic and doxastic changes that occur to agents after a promise is successfully made.

Update Semantics is a formal framework with the slogan "You know the meaning of a sentence if you know the conditions under which it is true". The presented approach, Promissory Update Semantics (PUS) extends this to a multi-agent, doxastic, and deontic setting, in the style of Dynamic Epistemic Logic. A series of promissory puzzles, which exemplify the differences between the promissory frameworks, are formalised and discussed in this dynamic setting.

Acknowledgements

I would like to thank my thesis supervisor, Prof Dr Frank Veltman, for his encouragement, insight, and feedback throughout this project, especially the hours of transatlantic skypeing involved in the latter weeks of writing.

Special thanks must go to my friends, family, and all those whom I have encountered along the way whom have been subjected to the odd rant or impromptu quiz on their opinions on this promise or that.

My fascination with promising was kindled throughout my philosophical studies in my undergraduate career at the University of Calgary. Between coursework with Dr Nicole Wyatt and Allen Habib, and articles written by other members of the department, I was subtly steeped in the puzzle that is promising.

The top of any list of thanks must go to my parents, Amin and Zarina Kurji, for their continued support in my pursuits of philosophy, logic, and life in general.

Contents

Chapter 1

Introduction

The act of promising has long been a problem discussed by moral philosophers, psychologists, sociologists, economists and many others. There are many views on what makes a promise and what promises themselves entail.

> *When setting up a time to meet a friend at a teashop it was brought up that I was habitually late for appointments with her. I responded with the following utterance "**I promise I won't be late for tea.**" As it happened traffic was quite heavy that afternoon and I was running late. Seemingly without contemplation my driving style became more aggressive as I battled towards being on time, and keeping my promise.*

While I made it on time for my appointment, I was slightly distracted by the ease in which I slipped away from obeying the traffic rules, though not dangerously so, in order to ensure I did not break the promise. Clearly there was a normative balancing act playing out during my commute. This started a fascination with *promising*: What is a promise? How do they work? Are there times when they are not binding? Do promises work the same in all cultures, i.e. are they societal institutions, or something more universal?

This interest in promising was fed by several courses, in philosophy of language and ethics, taken during my undergraduate days at the University of Calgary. The influence of my professors in Calgary, and their expertise on the topic, can be easily observed from a perusal of the bibliography of this work.

The aim of this thesis is two-fold. Firstly it should act as a survey of the major theoretical frameworks that explain promising, the bulk of this will be in chapter two. A new theoretical framework will also be introduced in that chapter, which avoids the pitfalls the current offerings suffer from. The third chapter continues this exploration via the contrast of promising with other obligation-creating speech acts.

Secondly a logical system, Promissory Update Semantics (PUS)[1], will be developed which can model the relevant factors which these frameworks require, in

[1] An unfortunate acronym to be sure.

chapter four. The aim of PUS is to be agnostic towards the various frameworks, we can model promissory situations under any of the discussed frameworks. That is, whether an example counts as a promise, or is outside of the promissory family depends on the philosophical view one takes, but the modelling will reflect all relevant phenomena.

This system will be applied to a series of 'puzzles' which exemplify the differences between promising under the various frameworks, in chapter five. These puzzles represent situations where a promise is considered to have occurred under one framework, but not under all of them. Here the technical formalism aids in showing just why it is the case that, for example, an expectationalist has trouble with a deathbed promise.

The final chapter contains discussions of the gaps that PUS leaves open, as well as other areas in which the update semantics can be applied towards modelling more complex puzzles

This paper is aimed at both technical and philosophical audiences, such that those without interest in the technical need not avoid it. Generally we will start every section with philosophical discussion and conclude with the technical modelling. In this way, the more technical (or philosophical) of the audience can gain a fruitful read without having to hunt and peck throughout the text.

Chapter 2

What a Promise is: Fixing the Concept

> To promise someone to do something is to commit oneself to that person to do that thing. [29, 287]

Asking a person on the street will yield a similar enough definition, though perhaps less dry in vocabulary. While practical, such definitions do not shed light on the nature of promises nor when a promise could be said to be made, or not, under what circumstances.

Before progressing much further we must be clear on what we mean by a promise. The prototypical promise is an utterance from one person to another wherein the words 'I promise' occur, e.g. the utterance "**I promise to pick you up at the airport**".

This is does not exhaust the set of promissory acts. Consider the following situations.

- A friend asks you "Will you do this huge favour for me?" **You nod.**
- Unable to pay your entire bar-tab you **write an IOU** for the remaining sum.
- When discussing jobs to be done at your sister's wedding you say "**I will definitely do that.**"

Taking the loose definition from the top of the chapter these too seem as if legitimate instances of promising, or at least one needs to say more in order to disallow them from the category.

2.1 Assertions and Intentions

It is clear that a promise is a special kind of utterance. The operative verb in our working definition is *commit*, and it will serve well to mark the difference between a promise and an assertion.

> An assertion is a speech act in which something is claimed to hold, e.g. *that there are infinitely many prime numbers*. [22]

Assertions are truth-evaluative, they can either be true or false. Put another way, an assertion rests in *facts*. A promise is of a different type than assertions. We promise actions that we can perform. They are *active* not factive.

I cannot **promise you that the grass is green**, but I can *assert* that this is the case.

We must be careful of a certain linguistic trap involved. The verb 'to promise' can be used for factive sentences. This does not necessarily mean that a promise is made. In English these sentences are viewed as grammatically correct. However this is not the case inter-linguistically.

In the Dutch language, for example, sentences like "**I promise that I will pick you up at the airport**", the employed verb is *beloven*. For sentences like "I promise you that the grass is green", using the verb *beloven* is ungrammatical, one must instead use the verb *verzekeren*, which roughly translates as 'to assure'. The former verb is associated with active utterances which commit the speaker, while the latter of factive utterances.

If promises are about actions, perhaps they are simply a variety of *statements of intent*. Statements of intent are statements that describe future actions to be undertaken by the speaker. The question is, is there is a difference between:

i. **I will pick you up at the airport.**

ii. **I promise that I will pick you up at the airport.**

A promise is not merely a descriptive utterance, it imposes moral a obligation. While one may expect to be picked up at the airport, in (i) they do not have a moral claim if this does not occur. The promise of (ii) brings with it precisely this moral factor. A promise is an utterance which describes an *obligated* future action on behalf of the promisor to the promisee.

> A promise is not merely an expression of intention. It is the assumption of an obligation. Promising obligates: that is the point. [9, 30]

2.2 Obligation-creating Circumstances

In order for an utterance to count as a successful promise it must occur in *obligation-creating-circumstances*. We have seen that it must be clearly more than an assertion, or a statement of intention; that it must be active and obligating. The various accounts of promising differ on what constitutes such circumstances, but there are two requirements that are common to all frameworks: A successful promise must pass the tests of *ability* and *preference*.

2.2.1 Promisor's Ability

> *Walking down the street you encounter a hobo. They say to you "If you give me €10 today,* ***I promise that I will give you €1,000,000 tomorrow****".*

Most would agree that this is situation does not constitute a legitimate promise. However it does match our requirements thus far. The utterance is unambiguous[1], and predicates a future action on the part of the promisor. What is missing? The belief in the ability of the promisor to deliver the promised act.

Without this belief the promise fails to have uptake. Few, if any, would say that if the homeless person was found the next day and they did not provide one million Euros that they could be counted as a promise-breaker. In order for a promise to go through, to have uptake, there must be a belief by the promisee that the promisor is able to perform the promised act. Note the distinction that the requirement is not that the promisor has the ability to do the act, just that the promisee believes that this is the case. If the promisor does not have the ability to perform the act, but the promisee believes that they in fact do the utterance is still considered a promise. This would be an infelicitous promise, in the same vein as the promise made but where the promisor does not intend to keep. The promise is still made, though under ethically dubious motivations, and is considered *broken* if the act does not occur, regardless of the (hidden) initial intent of the promisor.

2.2.2 Promisee's Preference

> *A mother, frustrated with her child not picking up after herself, says "If you don't clean your room,* ***I promise I will cancel our internet service.****"*

Here we see the words 'I promise' are uttered, it is *active*, it looks to obligate the mother to cancel the internet in the case of the room not being cleaned. Lastly, it is in the mother's power to cancel the internet service.

The point of contention here is that the promised act is not one in which the promisee (the child) wants to be performed. The 'promise' by the mother appears to be an inducement for her child to do the opposite of the antecedent act.

Although the word *promise* is used, this speech act is a veiled threat: an inducement to do a certain act on pain of some punitive action by the speaker. Here we see an example of the speech act and the utterance 'promise' not being coextensive. In order to be in obligation-creating-circumstances the promisee must not be against the performance of the promised act. One of the unique facets of promissory obligation is the ability to waive the obligation by the promisee. At any point between the moment the promise was made and the performance of the promised act the promisee can unilaterally absolve, or cancel, the promise and therefore the obligation of the promisor to perform. As

[1] We assume a non-sarcastic presentation implicitly.

such, even if one were to assume uptake on a negatively preferred act would immediately result in absolvement, ostensibly leaving the world in a situation as if the promise did not take place.

Not all promissory frameworks require pure unadulterated *want* by the promisee, some only need that they are neutral to the act taking place.

> *Your lawn is overgrown, but you are indifferent as to whether you mow it, or someone else takes on the task. Your neighbour Tom* ***promises to mow your lawn after he has dealt with his own****.*

In this case it would seem that, despite your indifference, one would not claim that a promise had not, or could not, take place, assuming even tacit ascension.

However if you were *opposed* to others mowing your lawn, because of, say, your exacting standards or the catharsis of the act, then the promise would not have uptake as it would surely be rejected.

2.3 Promissory Frameworks

Promissory frameworks generally fall into two categories. The first, *conventional* accounts, are based in contractual analogies and appeal to social convention. It is the convention that gives a promise its force and all obligations that result from a promise come from the convention as well. The second are *expectational* accounts. These theories centre on the expectations that are generated by the promisee as the key to promissory obligations. The force of a promise, and thus the injury when a promise is broken, is based in these expectations of the promisee and nothing else. A promise is not said to have been created unless expectations have arisen from the statement.

More recently theorists, such as R.S. Downie, have proposed hybrid accounts which marry the two together. What it is to be a promise is formed by both convention and the expectations formed in the promisee. We present general versions of then conventional and expectational accounts, followed by their main weaknesses. Then we will present our own hybrid account, influenced by the hybrid approach of Kolodny and Wallace. [18] The chapter concludes with a discussion of the speech act approach to promising, and how it can aid in the modelling of the promissory phenomena, without biasing towards a single type of framework.

Any theory of promising must somehow account for the fact that when a speech act, equivalent to "**I promise that** x", is uttered a moral obligation is formed for the speaker to perform x.

2.3.1 Conventional Accounts

The conventional accounts of promising hold that promising is based in a convention, a set of rules and practices, which encourages community and individual participation in the convention by allowing cooperation and coordination through trust. [16]

David Hume was one of the earliest proponents of the conventional account of promises. He details the conventional account of promises in the *Treatise of Human Nature*. There are other flavours of conventional frameworks, however we will stick to Hume's as a characterization of the family. At their core the conventional accounts have promising as a human, thus artificial, convention; putting promises as a type of *contract*.[2] That is, there is no *natural* reason for us to keep our promises.

> That the rule of morality, which enjoins the performance of promises, is not natural, will sufficiently appear from these two propositions, which I proceed to prove, viz. that a promise would not be intelligible, before human conventions had established it; and that even if it were intelligible, it would not be attended with any moral obligation. [17, 516]

The morality of a promise, then, stems from the communal acceptance, and practice of, the institution of promising. That is, the normativity of promising is directly tied to sympathy with the public interest.

Hume's conventional account of promising involves the following assumptions:[3]

1. Promising must involve an artificial, quasi-legal institution.
2. Promises are invoked by an indicating utterance.
3. Promises are not valid if made under duress
4. Promises are distinct from assertions, intentions, and making resolutions.
5. There is mutual advantage to both parties.
6. The promisee can waive the right to the promise, which completely absolves the promisor from the obligation.
7. The right to performance is forfeited under certain circumstances.
8. The promised act must always be wanted by the promisee.
9. There is always a promisee.

These points make up the core of an artificial institution created by a community that is beneficial at both the individual and community levels.

[2] Some have 'promise' as the normative foundation of the contract and legal systems.

[3] These are directly borrowed from R. S. Downie's survey in [11, 260].

Problems

One of the chief complaints waged against the conventionalist is the question of how the practice could have evolved. In order for a practice of promising to gain traction there must be some practice that deals with creating binding obligations between parties; this runs a serious risk of circularity.

T. M. Scanlon brings up two strong problems with the conventionalist approach. First note that the normative force of promising, that is what grounds the act in the normative versus the descriptive, is tied up with the convention. As such when a promise is broken the wrong is against the *convention*, not the promisee. The promisee is wronged as a member of the community who have volunteered to participate in the convention, the harm that applies to all members of the community. Following this any member of the community would have the same rights from a breach of promise as the promisee. [16]
Scanlon's second argument involves the following 'state-of-nature' case.

> *Suppose I am stranded in a strange land. In an attempt to get myself something to eat, I make a spear. I am not very good at using it, however, and when I hurl it at a deer it goes wide of the mark and sails across a narrow but fast-running river. As I stand there gazing forlornly at my spear, lodged on the opposite bank, a boomerang comes sailing across and lands near me. Soon a strange person appears on the opposite bank, picks up my spear, and looks around in a puzzled way, evidently searching for the boomerang. It now occurs to me that I might regain my spear without getting wet by getting this person to believe that if he throws my spear across the river I will return his boomerang.* [25, 297]

Scanlon posits that this scenario has the same essential ingredients as needed for a promise. Our main ingredients, thus far, match the scenario well, the only trouble being the lack of a direct utterance, not a necessary condition as we have already seen. He continues,

> *Suppose that I am successful in this: I get him to form this belief; he returns the spear; and I walk off into the woods with it, leaving the boomerang where it fell.* [25, 297]

This leads him to the conclusion that the wrong is no different than if he had promised to return the boomerang. Thus, as there was no social practice at work promising must be of a different variety.

To accept this argument we must accept the premise that the wrong in breaking a promise is the same as the wrong in abandoning the boomerang. It is uncontentious that Scanlon has betrayed the trust of the boomerang-wielding stranger. Not all breaches of trust are the same as a broken promise, however.

If I asserted that I would pick you up at the airport, and I failed to do so the harm is not the same as if I had promised to do so. I have disappointed you, or let you down, but I have not broken a promissory obligation. This disappointment

could be planted in a trust that you were to do so, without it becoming a full-fledged promise. Another way of reading this would be that the non-verbal communications were *statements of intent*. That is, though surprising that he walked away without dealing with his part of the tacit bargain, Scanlon is guilty of assertive deception and not of promise-breaking. Scanlon must do more, and explain why we should treat the weapons-on-the-river-bank the same as a broken promise. Hume's fifth criterion, *there is mutual advantage to both parties*, serves up the problem of *no selfless promises*.

> *You notice your elderly neighbour's lawn is overgrown. Feeling the good Samaritan you* ***promise to mow her lawn that afternoon*** *despite the scorching weather, and any form of restitution.*

The staunch conventionalist must not consider such a situation as promissory, as there is no benefit for the promisor to fulfil the act.[4] There are varieties of conventional promising which weaken this requirement, but they must say a more complicated story on how the convention is held up, if not by universal utilitarian means.

2.3.2 Expectational Accounts

The expectational accounts of promising hold that a promise is grounded in the resulting beliefs (expectations) of the promisee after a promise has successfully taken place. The normative force of a promise is entirely bound up in the reliant expectations of the promisee. There is no need for convention.

Promises are the sort of thing that invite *trust* that the act will occur. This trust is important, and when a promise is broken it is this trust that suffers. The normative force of a promise is bound in the (potential) harm to the promisee that stems from relying on such expectations. [16]

Just as Hume is representative of the conventional accounts the expectational accounts have their roots in the work of Adam Smith. Scanlon moved from the conventional approach to an expectational approach, after considering the problems discussed above.

> A promise is a declaration of your desire that the person for whom you promise should depend on you for the performance of it. Of consequence the promise produces an obligation, and the breach of it is an injury. [30, 472][5]

The obligation of a promise stems from the dependence of the promisee, and thus the harm from breach is one of deception. We must note that there are two distinct breeds of expectationalist. There are those who claim that, as the belief must be *reliant*, the potential breach must involve *tangible harm* in order for a

[4] We ignore, for simplicity's sake, the notion of residual benefits from feeling good about one's actions.

[5] via [11, 263].

promise to have been made. Others have the harm at the much lower threshold of *disappointment*.

Smith differentiates between the two approaches of promising in that the conventional accounts are concerned with *jurisprudence*, or the law, which is often conflated with *casuistry*, or ethics. The former is solely concerned with what an impartial observer would deem the promisee is entitled to exact by force. The latter, on the other hand, is concerned with the rules for the conduct of a good man. [11, 261] The expectationalist accounts live in the space of ethics, not the law.

Duress

A key difference between the conventional and expectational accounts is their treatment of *promises made under duress*.

> *A highwayman corners a traveller and, disappointed with the amount of money the traveller had to give, extracts a promise, on fear of death,* ***to provide the highwayman tens times the amount the following week.***

For Hume it is clear that this does not a promise make, flying directly against criterion (3) and clearly not in the spirit of (5) either. From the view of the law it is clear that there is no obligation formed; this is usually directly encoded in the legal systems. For Smith's purposes this is not as clear.

Expectationalist accounts have the full normative force of a promise rest with the resulting beliefs in the promisee, in this case the highwayman. If the traveller convincingly acquiesced, promising the tenfold return, then, if he be a good man, he could feel bound to follow through with the promised act.

> Whenever such promises are violated, though for the most necessary reasons, it is always with some degree of dishonour to the person who made them... Fidelity is so necessary a virtue, that we apprehend it in general to be due even to those to whom nothing else is due, and whom we think it lawful to kill and destroy. [30, 332]

Problems

The principle criticism of expectational accounts is that by only grounding a promise in expectations it reduces promises to the same sort of act as *advising*, *warning* and *threatening*. [16] If there is nothing more to a promise than an expecting belief then the difference between a statement of intent and a promise would be a matter of degree. With this the distinction is in danger of sailing on Theseus' ship.

Recall that the difference between intention and promise was the formation of an obligation. When we lose the ability to discern between the two, we also lose the ability to explain how a promise produces an obligation, while stating an intention does not. It is not clear what, if anything, a promissory obligation

is under this reading. Breaking a promise is akin to common deception, which also occurs with intentions and assertions.

A promise is a speech act that indicates to the hearer 'rely on me!' — it is a linguistic convention to indicate speaking in earnest.

2.3.3 Hybrid Accounts

In recent years there have been attempts to reconcile the difficulties of both types of system by marrying them together, creating hybrid accounts. Adam Smith presented the two frameworks as differing *points of view*. This, then, is not a question of rightness or wrongness but of theoretical aim.

If there is a conventional component to promising, then a discussion of breach should include elements of said convention. On the flip-side, as promises are about commitments and obligations, not addressing the beliefs of those committed to seems equally misdirected. Hybrid accounts of promising take both of these points to the explanatory heart.

Some accounts are *reductionist* in nature, reducing promises into membership to a larger category of obligation-creating speech act. An example of such a move is the approach of R. S. Downie. He posits that promises, in their nature, are a form of *pledge* and so their main normative force derives from a form of *self-fidelity*. However added to that is the fact that promising is clearly a form of social practice, and as such when they are made the promisee acquires additional rights against the promisor. [11, 269]

In this way the account is *hybrid* we have aspects of convention, as well as expectation. These added rights, of the promisee, are both lesser and not necessary for the promise to take place. They are not the morally fundamental reason for one to keep their promises.

Hybrid accounts address the convention-expectation dichotomy by describing multiple layers of obligation. As we saw with Downie, the result is a more fine-grained tug-of-war between the two factors. Ultimately we could say that Downie is a conventionalist with a flavour for expectations. This is because he acknowledges the normative force of expectations, but still holds that one can have a promise without dealing with such factors.

The later Scanlon takes the opposite approach; promises are primarily about expectations. Even if there is a convention in place, there need not be in order to explain the phenomenon. Scanlon stays completely in the expectationalist camp because of this non-necessary view on the convention.

Kolodny and Walace, in *Promises and Practices Revisited*, challenge Scanlon's expectationalist view. They ultimately conclude that we cannot ignore the one side for the other.

> [A] satisfactory theory of the obligation to keep promises must make reference to the mora implications of the fact that promising is a practice. Only by doing so can we capture the distinctive way in which uttering the formula, I hereby promise to do X, gives rise to

> a moral obligation of fulfilment under the more general principle of fidelity.

Their discussion falls short of proposing their own framework of promising, however. We will take the path started by Kolodny and Wallace, and present a hybrid promissory framework.

The Proposed Account

The core of Kolodny and Wallace's account was to take on Scanon's expectational viewpoint and show that it was lacking unless the conventional nature of promising was also dealt with. A successful account of promising must not privilege one side over the other.

Unlike Downie's account we also do not want to present the two sides as separate, i.e. divorcable, normative parts. It is not enough to call the two sides 'equal' in normative standing, and continue along Downie's way.

Instead we argue that they are *intertwined*; that the nature of promising is one of *conventional expectations*.

With Hume we saw that there are a number of rules to promising, specific situations which reflect obligation-creating-circumstances. From Smith we get that the normativity of promises has to be about the conduct of 'a good man', i.e. it must not be about the provably faulted but doing right by the promisee.

We propose that these are, as alluded by Smith, two sides of the same coin. The good man must do right by others, but how he does right (in this instance) is governed by the social convention that he is partaking in. While a social convention can described a moral practice, it cannot create moral force from thin air. This is precisely why legal obligations are so tightly defined in practice, and of a different *type* than moral obligations. An oral contract is a different thing than a promise.

For a strict contractualist, there is no natural motivation for promises: the obligation is purely artificial. What is missing from Hume's account is from where does promises gain its *obligatoriness*? If it is just by participation in the convention, then one could surely opt out, or in some other way signify non-participation. However, not keeping a promise under the claim that one 'doesn't promise' goes against ones intuitions of what it is promising is. We solve this dilemma by appealing to the expectations of the promisee as the normative grounding for the practice. In this way we cannot split the two apart. The convention of promising is *normatively based* on the incurred expectations of the promisee. Similarly, the expectations are guaranteed to be generated because of the details of the social convention involved.

This approach avoids Scanlon's problem of needing the practice: the two men could have convinced each other into reliant expectations, while not refuting the nature of promising as extra-conventional. Hume's problem of opting out is similarly avoided, as we can incur reliant expectations outside of the convention. The social convention of promising, however, is an explicit prac-

tice which governs the appropriate creation of such beliefs, as well as provides rules in which the resulting obligation may be waived. We will call this the *conventional expectations* theory of promising.

2.4 How Promises Work: Speech Act Theory

While the various accounts differ on what is needed for, and followed by, a promise there is a core that is similar throughout. Promises have also been dealt with, famously, in speech act theory. We will discuss this approach to suss out our strategy to develop the formalism, keeping the factors needed by the philosophical approaches. We will show that all three promissory frameworks can be described in terms of speech act theory, and this normalisation will serve as the foundation to which we will build our formalism.

J. L. Austin introduced speech act theory and with it the study of promising gained a new approach. Austin introduced the notion of the *performative utterance*, an utterance which itself is an action. Performatives change the world, they don't merely describe it. He defines the performative as sentences in which "to utter the sentence (in, of course, the appropriate circumstances) is not to describe my doing of what I should be said in so uttering to be doing, or to state that I am doing it: it is to do it." [2, 6]

The speech act of promising was one of the standard examples of the performative. The statement "**I promise to pick you up at the airport**" is not a description of a promise being made, the statement itself *is the act of promising*. That is, there is no other thing that is the making of the promise to which the utterance is describing. Performatives are active, rather than factive, thus they can be neither true nor false. The promise to 'pick you up at the airport' is what creates the obligation to do so.

Austin developed the idea of the performative into a triumvirate of speech acts: the locutionary, illocutionary and perlocutionary. In our example promise, the *locutionary* act is the actual utterance of the sentence "I promise to pick you up at the airport". The *illocutionary* act is the active element of the utterance, the change to the world that ensures the utterance is not purely descriptive, here it is the creation of the promise itself. Lastly there is the *perlocutionary* act, this is the psychological consequences that result from the act. Not all utterances contain a perlocutionary act, though when there is one it is often in the listener.

The speech act approach to promising invokes a 'convention' of sorts, so is often paired with the *conventional* accounts of promising as well. However, the convention that is needed for the speech act approach need not be societal, it can be a linguistic convention. Part of all promissory frameworks is the need for the promise to take place in obligation-creating-circumstances. That is, there is an underlying practice, set of rules, that indicate when a promise is taking place and when one is not. True some varieties of expectationalist may blur the line between statements of intent and promise, but there is some foundational difference which *obligates*. How and when this takes place is the 'convention'

that is needed, and speech act theory calls for nothing beyond such a foundation.

J.S. Searle who also employed the example of promising in his discussion of the mechanism of the illocutionary. The illocutionary act of promising, according to Searle, is a matter of convention.

> For the case of promising and statements there must be some conventional elements the utterance of which counts as an undertaking of an obligation or the commitment to the existence of some state of affairs in order for it to be possible to perform such speech acts as promising. [27, 40]

The speech act approach to promising ties the parts of the performative utterance to the required elements of the promissory frameworks. The obligation of a promise rests with the promisor, while the belief with promisee. Smith's discussion of the legal and ethical points of view reflect this difference as well. Searle is ultimately a conventionalist. So the normative nature of promising is wrapped up in the *illocutionary*. Austin's discussion of promising paints him as an expectationalist of sorts, as the *perlocutionary* takes precedence in his view.

With this interpretation of the nature of promissory speech acts we can analyse promises by their illocutionary and perlocutionary effects, which maps to the conventional and expectational approaches as well. The conventional expectations, or simply 'hybrid', approach can be shown by modelling both the illocutionary and perlocutionary effects.

Chapter 3

What a Promise is Not: Refining the Concept

With our promissory frameworks in hand we have a handle of what it is that makes a promise a promise. However, promises exist within a family of obligation-creating speech acts which are often conflated with one another, or at least treated as deontically similar in study. This chapter will explore these related, but distinct, phenomena as juxtaposition to further clarify what a promise is. We begin with the question of to whom can promises be made.

3.1 Promises to the Self

One of the areas of contention, surrounding promises, is whether one can make a promise to the self. That is, if the promisor is identical to the promisee does this constitute the correct type of circumstance for a promise to form. While it is clear that people can obligate themselves to themselves, it is not yet clear that this ability extends to promissory obligations.

> *Having just moved to Bristol, a hilly city, from Amsterdam, a topographically uniform city, you are surprised at how out of breath the fifteen minute commute to work gets you. "**I promise that I will exercise regularly and get fit, such that I am not a sweaty mess at department meetings.**"*

Here is an example of a promise that fits with our various promissory frameworks, with the exception of the operative question at hand. The promisor has the ability to get fit, the promisee clearly desires the act to occur, etc. Should we then consider this a promise?

Allen Habib is one of the strongest proponents of the promise to the self being a legitimate form of promise. His authority theory holds that promises have normative force because they are commands we give ourselves, as authorities over ourselves. The thrust of his argument is precisely that traditional accounts

of promising fail to explain how we can make promises to ourselves. We will deal with the differences between commands and promises in section 3.3, but let us address potential problems with promises to the self, whether command-driven or otherwise.

Promises are norm-creating acts, that is new normative restrictions are placed on the promisor after a promise is successfully made. Unlike strong unchanging morals, e.g. the Ten Commandments[1], promises are not forever binding. Built-in to the convention, linguistic or institutional, are the ideas of waiving and absolvement. At any point between the promise being made and the act not yet performed the promisee retains the right to waive the obligation towards them, in (normative) effect erasing the promise. Similarly, if circumstances change drastically one is not held responsible for the performance of the promised act if it no longer is under their control.[2] Release from, or the waiving of, a promise is done solely by the promisee. Absolvement is usually characterized by a known, by both parties, change in circumstances, but it may also be petitioned for by the promisor. We add to the airport example that my boss has made me come in on my day off, when you are to arrive, and I will have to use my entire lunch-hour rushing to pick you up and drop you off. While I can still follow through on the promise, it would be to a much greater inconvenience on my part, given your neutrality to riding on trains. The decision remains with the promisee, but it can be triggered or requested by the promisor.

The idea of unilateral release of a promise is the central problem for any account which allows promises to the self, *agent-reflexive obligations*.

In normal description one can be released from the obligation of a promise when the promisee allows for it to be waived. However, in these cases, the promisee and promisor are one and the same, so it would seem that one can always get out of the promise via self-release. In this way the speaker is never truly *bound* by the obligation to which they have entered, without having to *petition* the promisee for absolvement.

The idea that a promise is not binding, insofar as the speaker can simple cancel it and move on, is quite troubling and needs to be addressed by any accounts allowing for such matters. This gap could be patched for this sub-class of problems, e.g. the waiving of obligations of this type does not totally rescind the obligation but removes the negative repercussion of breach. Such approaches then split promises into, at least, two very different types of obligations; the immediately defeasible, and the externally binding.

Recall the difference between a promise and a statement of intention was precisely the binding obligation. Here there is no practical difference between a promise to the self and a statement of intention, if it is treated in a similar manner to promises.

The promise to the self, like the above example, bears strong resemblance to other types of normative acts, e.g. New Year's resolutions. Resolutions,

[1] Written in stone, literally and metaphorically as they are.

[2] Consider the promise to pick you up at the airport, and the theft of the promisor's car. One would be hard-pressed to call this a substantive breach of the promise, unless the promisor actively solicited the theft in some strong way.

like pledges, operate under different normative force than promises. Breaking a new year's resolution is still considered a breach, but one cannot merely absolve themselves of their resolution to avoid the breach. Practically promises to the self are more in line with these phenomena than standard promises. While the word 'promise' may be used, it seems more likely that this is a case of borrowing the normative strength for emphasis, rather than a legitimate instance of promising. A further discussion of pledges can be found in 3.4.

It is clear that the normative force of promises necessitates that the obligation is between two distinct parties. This objection lead to a rejection of promises to the self, and thus are a problem for any reductionist approach to promises as reflexive-commands.

3.2 Promises to the Many

On the face of it promises to the many seem unproblematic, but as we saw with the self, the many too have unexpected difficulties.

Recall that a promise requires some form of uptake.

Let us think smaller first.

> *A mother, to her two darling children, says "**I promise that we will go to your favourite restaurant this weekend**".*

There are multiple ways that this can be interpreted. We can view this as two separate promises stemming from the same utterance. This interpretation would especially make sense if the two children have different favourite restaurants. The other way to interpret this situation is as a *single promise*.

Let us assume that both children share tastes in culinary destination. Consider the situation where the mother takes one, but not both, of the children to said restaurant on the weekend. In the former interpretation we have a mother who kept one promise, and broke another. The latter we strike a problem, the promised act was not performed for a promisee, thus we have breach. If we asked the one child if their mother kept her promise, are they likely to say 'no'?

The first situation is two instances of promises to the one, and unproblematic. The second leave us puzzled, a promise to more than one can be kept, broken, or both. One could take the more strict approach and claim breach unless every 'branch' of the promise is successful this avoids the trap but it is not enough to escape the trap of the many.

As you increase the number of promisees the standard interpretation of a promise seems less applicable. When one makes a promise to the arbitrary many, e.g. a promise to society at large or 'the church going flock', we encounter the problem of *uptake*.

Recall that obligation-creating-circumstances require that the promisee(s) be at least neutral towards the promised act.

> *To five of your colleagues you "**promise to change the TPS report policies**". One of them, however, is adamantly attached to the current handling of the TPS reports, and is against any changes.*

If this is viewed as a single promise and you do change the policies the status of the promise seems to be held. Before that we have the issue of whether the promise took place. Not all of the promisee will have assented to the promise, and if there is no universal uptake then the promise should not have taken place. That is, we have problems with both preferences and uptake. [14, 71]

Promises directed at more than one seem best described as sets of individually aimed promises, especially when the performance of the promised act is not necessarily intertwined. Promises to the many, where the act is singular and the audience many, run into uptake problems. The problem of the many is not as strong as the problem of the few, but one must be careful in how they interpret such promises as the occurrence, and follow-through, of the promise varies depending on which interpretation is utilised.

3.3 Commands

Promises and commands are often treated similarly. They are both obligation-creating speech acts that occur between one and an other(s) agents. Because of this similarity they are often modelled in like systems, e.g. [38]. There are crucial differences between these two speech acts that should prevent such treatments.

On the surface the main difference is that commands and promises differ in the direction of obligation. With a promise the speaker becomes the obligator, and the hearer the obligatee. For commands it is precisely the reverse; the speaker commands and this obligates the hearer.

This makes them fundamentally different acts. A promisor is consciously choosing to be obligated and perform some act for the benefit of the promisee. A commandee has no choice but to perform the act when they become *aware* of the command.

The inverse remains through to absolvement. The commandee has no way to absolve, in the same away the promisee does. Moreover there is not a situation where the commandee, as the obligated, can appeal for absolvement in the same way the promisor may request.

This is because of the status differences between the active parties. A promise can take place between any two agents, their relationship does not matter. Commands require a pre-existing *authority relationship* in place before it can take place. The only way an agent can command another agent is if they have some sort of authority over them.

Promises involve a convention, whether institutional or linguistic, which dictates obligation-creating-circumstances. Commands do not rely on any convention, so long as a command is understood to take place, i.e. uptake is secured, the command will occur.

A convention could exist that grants the authority relationship, but the speech act itself would not rely on the convention, only its effect.

Commands look very similar to promises, being obligation-creating speech acts between two distinct individuals, but we know that they differ on their normative force. Obligations from commands are generated through *authority*,

and thus derive from a pre-existing relationship between the two individuals. Further the doxastic change that may, or may not, result is unimportant to the command, i.e. there is no connection between the obligation's force and the change in beliefs of any agent. We will call the speaker i, and the hearer j. Most commands are of the imperative form "j, see to it that φ". The commandee is expected to perform the act φ. In promises we see the reverse, they are of the form "i will see to it that φ". Even if we reversed the roles of the speaker and hearer to model commands, it would seem wholly out of place to ascribe the belief of the commander that j will φ as the, or part of, grounding of the obligation.

3.4 Oaths, Pledges and Vows

> [W]e should note that vows, pledges and oaths have the same (apparent) normative structure, and the same functional role, as promises. They are sets of words that, when uttered under the correct conditions, cause the speaker to become obligated to perform some action, which is in turn specified in the text of the speech. This is exactly the normative structure of promises. Further, functions of vows, pledges and oaths are the same as those of promises, i.e. to undertake commitment, to offer assurance, to demonstrate loyalty. [14, 42]

Ask the man on the street the difference between an *Oath*, *Pledge* and *Vow* and you are likely to encounter initial puzzlement.[3] On first blush it would seem that these phenomena are interchangeable.[4] The average dictionary yields a definition of all three phenomena as 'a solemn promise'. We will show that this is both true, and false, and show how these fit in the promissory framework.

3.4.1 Pledges

> **pledge, v.** a. trans. To become surety for, make oneself responsible for (a person, thing, or statement). Obs.[5]

To pledge is then to obligate oneself for various phenomena. One cannot pledge that p is true, rather that they are being honest in stating p. Pledges are about future *action*. We can thus leave the factive behind, but still ambiguity remains. The brevity of the definition shows that the obligatee of pledges is not a set thing. Following Habib we note that there are three types of pledges. *Donor pledges* are pledges of possessions to another person, such as those made to the underfunded public radio station. The remaining two are related by fidelity; *pledges of fealty* are those that espouse loyalty to some person, group,

[3] Indeed before contemplation the present author was troubled by this very question.

[4] An example of this type of conflation can be found at `http://www.spiralscouts.org/node/35`.

[5] "pledge, v.". OED Online. June 2012. Oxford University Press. 3 July 2012 `http://www.oed.com/view/Entry/145634?rskey=Dws3VC&result=2&isAdvanced=false`.

or institution, e.g. a 'Pledge of Allegiance' to a country, or its monarch. Finally we have *pledges of comportment* which are pledges to adopt certain standards of behaviour, e.g. a pledge of sobriety [14, 74-75]

Consider the *donor pledge* "I pledge €25 towards supporting your programming". during a radio station's annual funding drive. This is a speech act which results in both an obligation to perform a future act, on the part of the pledger, and a belief about the act's occurence, on the part of the pledgee. This seems on the side of our standard promises.

A *pledge of fealty*, again following Habib's lead, to a specific person seems uncontroversially a variant of our standard promise. The pledge of fealty to a country can be just that a promise to a country, or institution. Though it is unclear whether one can do that, as we cannot reserve uptake of the promise, let alone an expectationalist's need for belief change. We will return to this in a moment.

Pledges of comportment encounter similar difficulties. They list a series of behaviours, or activities, that together describe a standard. However, there is no direct mention of an obligatee, i.e. to who the wrong is attributed in the event that the pledge is not upheld. If its society as a whole, then we encounter the problem of promises of the many, however this seems unlikely as a general move. Habib notes that pledges of comportment are usually ameliorative, i.e. one makes such a pledge because they feel that the described behaviour is good, and not currently being performed (on the pain of vacuity). So we have that pledges are made 'for the sake of' the pledges; thus are made to oneself. [14, 79] The obligator and obligatee are one and the same.

These *agent-reflexive* obligations need not be public, as the uptake is guaranteed at the outset. However the irreflexive versions, at least in the donor case, need be communicated in much the same way as a promise is held to need.

Now, we may ask under which normative force do pledges operate. It cannot be authority, for if that were the case the pledge of fealty would be redundant, as it is the act to which authority is granted. For the comportment case, this seems to be an appeal to the honour of the pledger. If pledges can be agent-reflexive, then we run into issues of the promise to the self. If we take the obligation to be all of society then we have gone too far and it is clear that we will have most, if not all, of the problems of promising the many at play. We are left with the obligation resting for, and with, the utterer themselves.

In order to support this claim, consider when a pledge is breached. If it is a pledge of comportment, made publicly, the only thing another person may do is reveal disappointment in the breach. That is, the punitive reply to such pledges is reproach. In the case of sobriety pledges this maps well. When the alcoholic falls off the wagon his friends and loved-ones are awash with disappointment at the fall; and they can choose to express this, or not. However it is uncontroversial to claim that they are not morally wronged by this.

Let us return to the donor pledge, though we assumed it was a promise of the standard sort. Having worked (volunteered) at a campus radio station[6] for

[6]CJSW, Calgary's Independent Radio 90.9 fm `http:\\www.cjsw.com`.

a time that included two funding drives, I can relate personally to the handling of such matters.

> The funding drive takes place over exactly one week, thus allowing each program a chance to appeal to their listener base for support.[7] At the end of the week, if the pledge was still outstanding it became the job of a choice few volunteers to politely track down the pledger and remind them of their pledge. After sometime, I believe but cannot fully recall it being shy of a month, we stopped such attempts and the remainder became the 'list of deadbeats'.

From this we can see a noticeable difference between a promise and such a pledge. A promise brings with it some form of punitive response, in the case of irrevocable breach, or the ability to call in the promise, for cases of simple open-ended promises like those of promised money. If I "**promise you €25 next week**", you can ask for the money during the next week. My only available recourse is postponement, but at the end of the week I am stuck with either paying, or breaking the promise. With the pledge, this is not the case. Even if we see the rogue pledger in the hallway, we can approach and *remind* them of the pledge but cannot come out and demand the money in the same way. For pledges, one can impeach, not claim breach.

Pledges of allegiance, and other fealty pledges, are less clearly in line with honour-bound normative force.

3.4.2 Oaths

> **oath, n.**
> a. A solemn or formal declaration invoking God (or a god, or other object of reverence) as witness to the truth of a statement, or to the binding nature of a promise or undertaking; an act of making such a declaration. Also: the statement or promise made in such a declaration, or the words of such a statement. The making of the declaration was expressed in early use mainly by swear, and later also by make or take, as to swear an oath , to make (an) oath, etc. [8]

Firstly we see that an oath, on the surface, differs from a promise in that it is a noun, however this is a minor barrier as the active verb 'swear' is tightly associated with the noun and we will treat *oath* as synonymous with *swear an oath* for reasons of clarity.

The core of an oath is the invocation of a deity, or other reverent object, as " witness to the truth..., or to the binding nature of a promise." Two things follow from this description. Firstly that an oath can be factive, or it can be about future actions. We leave the factive type here, for the reasons outlined above. Secondly inherent to oaths is an appeal to authority, almost exclusively

[7] Biweekly shows split the time, or more often shared the timeslot.

[8] "oath, n.". OED Online. June 2012. Oxford University Press. 3 July 2012 `http://www.oed.com/view/Entry/129495?rskey=5LYnI4&result=1`.

a normative one. The obligation, regardless of what it is, is to the normative authority. This claim has been contested, notably by A. Habib [14, 81]. There he claims that oaths which concern actions for a third party cannot be read so easily. How is the other party to claim breach, unless the obligation is pointed to them. Thus promissory oaths attribute the *right of punishment* to the deity, and the *right of performance* to the (mortal) third party. Given the degree of reverence to the appealed authority, if performance is not acheived that the third party need only invoke the wrong to the authority in an act of shaming/confrontation. The wronged would likely appeal to the same authority when attempting to invoke, or rebuke the lack of, performance. The specificity of the oath being to the authority need not lead us to a complex split obligation.

There is a second type of oath to be considered. These are oaths that do away to the appeal to nomrative authority, henceforth referred to as a *secular oath* (oath_{sec}). As the modern world becomes secular one would assume that oaths would have all but disappeared. But we still see oaths in modern society, most commonly in institutions. In Western legal systems individuals are *sworn in* to the proceedings. The standard swearing in is done on the Christian Bible, thus fitting our definition of an oath. However, in the United States and United Kingdom, one can substitute this for an *affirmation* which carries the same legal weight.

More common are oaths_{sec} such as 'Oaths of Allegiance' performed during citizenship ceremonies, or the Hippocratic Oath for newly minted medical doctors. An example is the oath of allegiance which is sworn by members of parliament in the Netherlands.

> I swear (affirm) allegiance to the King, to the Statute for the Kingdom of the Netherlands, and to the Constitution. I swear (affirm) that I will faithfully perform the duties my office lays upon me. So help me God almighty! (This I declare and affirm)[9]

Here we see the use of affirmation as the replacement for oath. That is oath_{sec} is essentially an affirmation in both the legal and political arenas. An affirmation is read as a strong assertion of the validity of what is being said. We note that affirming is noticeably weaker, in the normative sense, than a promise. Though not all of the above points against assertion will stand up to formal affirmations it is clear that these are distinct phenomena from promising. The targeted normative audience of such an act, if any, is unclear and we encounter the problem of the *promise to the many*. The character of oath_{sec} is that they are formal, public affirmations almost exclusively to institutions, which may be regarded as *objects of reverence*.

Now we can see a relation between oath_{sec} and pledges. We note that the oath_{sec} usually occur within institutional contexts. The breach of them conjures up the same set of responses as pledges. The observer cannot claim harm, but can impugn the character of the one who broke their obligation. The doctor who

[9]Translated from: `http://www.st-ab.nl/wetten/0430_Wet_beediging_ministers_en_leden_Staten-Generaal.htm`.

stands idly by at a medical emergency is chastised, but does not owe the observer for breaking an obligation to them, at least not for reasons directly related to the taking of the Hippocratic oath. Secular oaths are wrapped in institution and public utterance. One wouldn't undertake a secular oath in private. Thus we can characterise oath_{sec} as a subset of pledge; they are institutional public pledges. We see that this characterisation puts them in a different family of obligation-forming speech acts as *oaths* because the normative force is internal facing, rather an appeal to an external authority.

3.4.3 Vows

> **vow, n.**
> a. A solemn promise made to God, or to any deity or saint, to perform some act, or make some gift or sacrifice, in return for some special favour; more generally, a solemn engagement, undertaking, or resolve, to achieve something or to act in a certain way.[10]
> **vow, v.**
> a. With subordinate clause (or equivalent). The subject of the subordinate clause may be different from that of the verb itself.[11]

Traditionally a vow was considered a covenant between the speaker and God, or other higher order normative authority, e.g. Monastic vows. This is the consensus view, stemming from St. Thomas Aquinas. They are viewed as a subset of oaths;the difference being that one can make an oath about (to) anything (anyone). For oaths the normative force is the appeal to the authority, but the target of the speech act need not involve the deity. A vow is between the speaker and the deity itself.

The most common example of 'vows', in the current era, is that of the wedding vows, those uttered during the wedding ceremony. They are also one of Austin's quintessential examples of the performative, or at least the act of marriage in the form of 'I do'.

Let us compare the following examples of commonly used, religious vows:

> **Wedding Vows 1**
> In the name of Jesus, I _____ take you, _____, to be my (husband/wife), to have and to hold, from this day forward, for better, for worse, for richer, for poorer, in sickness and in health, to love and to cherish, for as long as we both shall live. This is my solemn vow.

> **Wedding Vows 2**
> I, _____, take you _____, to be my wedded (husband/wife), to

[10] "vow, n.". OED Online. June 2012. Oxford University Press. 4 July 2012 `http://www.oed.com/view/Entry/224757?rskey=R4u62H&result=1&isAdvanced=false`.

[11] "vow, v.1". OED Online. June 2012. Oxford University Press. 4 July 2012 `http://www.oed.com/view/Entry/224758?rskey=R4u62H&result=2&isAdvanced=false`.

> have and to hold from this day forward, for better for worse, for richer for poorer, in sickness and in health, to love and to cherish, 'til death do us part: according to God's holy ordinance, and thereto I pledge you my love and faithfulness. [12]

Here we see that in the first example the act is performed 'in the name of' the deity, and it is described as a *vow*. We note that it is left implicit that the agreement is not with the underlined party, rather with the deity, only on the reading of the final sentence labelling the act. Compare this with the second example. Here we have a list of activities, appealed to by 'God's holy ordinance' but not invoking the deity. This matches our idea of *pledges of comportment*, and it is labelled as we expect.

Just like with oaths we see that vows are still employed in the secular world. With the deity removed from both sides of the equation it seems that these collapse into the family containing oath_{sec} and pledge.

In accommodating for the secular weddings, which still contain the recitation of marriage vows, we could appeal to the *institution of marriage* as the new object of the vows. Though this seems suspect in multiple ways, the argument from romance seems enough to motivate the search for a more suitable explanation.

Elizabeth Brake, in her "Is Divorce Promise-Breaking?", addresses this issue in a way commensurate with the general approach described here. While we will not reproduce her arguments in full, we will cover them briefly as they shed light on the nature of secular vows. The discussion is of unilateral divorce in secular love-based marriages.

The first move is to be clear about to which we are referring; the wedding results in two sorts of obligating agreements between the wedded parties. The legal relationship is well-defined and has 'escape hatches' in the form of legal recourse. This is separate from the wedding vows themselves and what they represent for the two parties. The issue at hand only concerns the latter so we divorce the pair and concentrate on the uttered vows.

The question at hand is, when one (unilaterally) divorces, however, both agreements are broken. If these vows are promises, why is it the case that the we do not morally condemn the divorcing party to the same extent we do to promise-breakers? Her response is the following:

1. One can't promise what one can't do.
2. One can't control love.

In this way the major parts of the marriage vows seem to not be of the obligation-forming sort, as we do not have the right sort of uptake conditions for them to follow. Then are we to believe that marriage vows are meaningless?

Brake answers this concern by appealing to the non-emotional content of the individual vows spoken. These, taken in aggregate, form the set of conditions

[12] `http://christianity.about.com/od/christianweddingelements/qt/weddingvows.htm`.

that define a role: the spouse. The vows are a promise, or set of promises, to fulfil these conditions to the other wedded party. This formulation matches our standard definition of what a promise is.

Interestingly, Brake notes, the promise(s) in the vows can be broken at any time, regardless of the legal status of the relationship. "What couples promise in the end depends on what they say. And this suggests a mundane piece of advice—be careful what you promise!" [9, 38]

While this explanation fits with our current conception of promissory acts, it is not wholly satisfactory. It retains obligations between the parties, but at the expense of all emotional content. With a slight extension, I believe, we can keep the emotions.

If we treat the marriage vows as a pledge, in the same spirit of a *pledge of comportment*, we need not worry about the obligation being promissory, and thus fitting our strict requirements. The *role* being espoused is akin to that of the sober: a list of behaviours that are viewed as good in which the pledger will endeavour to pursue.

The more cynical in the audience could take this to the extreme. Those concerned with the frivolous and metaphorical nature of vows, or even the entire institution of marriage, could consider moving entirely to the pledge, thus not generating obligation in anyone but the speakers themselves. I prefer to think that wedding vows have evolved, via the changes in history and the move to secularity, such that they are more complicated phenomena involving both a pledge and a promise. However, both options remain available.

Vows are to god, we still call them that, but in light of todays secular society this seems to be more of a traditional label than an indicator of the current phenomena. The prevalence of writing ones own vows is good support of this, as it becomes increasingly unclear as to whether the religion at hand would prescribe all elements of the list. We can explain this by differentiating between vow and vow_{sec} which, much like oaths, can be seen as just another part of the family of pledges.

Thus we have,

$$\text{oath}_{sec} \subset \text{vow}_{sec} \subset \text{pledge}$$

Vows, oaths and pledges are similar as obligation forming speech acts, often they are bundled up with promises themselves. They differ from where they draw their normative force. With the increase in secular usage we see that the common conflation of the terms is warranted, or at least unsurprising, as the secular forms are virtually indistinguishable from each other save for labelling.

While most theories of promising do not differentiate between these phenomena and our standard promising, we would be remiss not to note these differences. Accepting Habib's argument that the pledge subfamily are actually self-obligating phenomena, given that most theories prohibit self-promises, would cause these acts to fall out of the promissory landscape.

Chapter 4

Building the Formalism

With this knowledge of the intricacies of, and differences between, promissory frameworks we now turn to formalising these notions into a logical system.

First let us restate the desiderata of what is needed to model the promissory frameworks. There will be items that are needed by some, but not all, promissory frameworks discussed in Chapter two. As we want a flexible, or agnostic, formal system all such factors need to be addressed within the formal system.

Definition 1. *(Desiderata)*

1. There is a promisor and a (distinct) promisee.
2. Promises can be spoken or written.
3. Promises are not factive, but active.
4. The promised act is in the future.
5. The promisor is able to perform the promised act.
6. The promisee is not against the performance of the promised act.
7. A promise results in a change of the beliefs of the promisee, usually this includes reliance that the act will occur.
8. A promise results in the obligation of the promisor to perform the promised act.
9. There are different notions of normative force.

At the core, promises change the information states of both the promisee and promisor. Deontically, promises change the normative state, they create obligations; promises are prescriptive not descriptive obligations.

There are several candidate approaches in modelling promises, each with its own shortcomings. Recently Yamada has extended the Dynamic Epistemic Upgrade Logic of van Benthem and Liu with parts of his own $MDL^{+}II$, to

model the obligation-creating speech acts of commands and promises [32], [37]. Promises and commands are treated close to identical in the resulting system, which given the discussion in 3.3 should not be the case.

We will be developing an Update Semantics to model promising, in the various frameworks discussed in chapter two. We will start with the Deontic Update Semantics (DUS) of van der Torre and Tan [33], which is based on the Update Semantics of Veltman (1996). Update semantics offers clear insight into how agent states change when statements are made to them, through the use of dynamic operations. In DUS obligations are actions which fits well with our speech act treatment of promises.

DUS is deontic, and partially epistemic, so we must thoroughly extend the system to obtain our complete desiderata. We will extend with deontic parts of Yamada's system and standard doxastic elements from the DEL family of dynamic logics, as well as elements unique to this system.

Veltman introduces his update semantics with the following slogan:

> You know the meaning of a sentence if you know the change it brings about in the information state of anyone who accepts the news conveyed by it. [35, 222]

With the addition of deontic states in their system van der Torre and Tan offer the additional slogan

> You know the meaning of a normative sentence if you know the change it brings about in the ideality relation of anyone the news conveyed by the norm applies to. [33, 4]

We continue down this path, though we must add more than just a deontic parameter to our update semantics. We will first discuss the basics of Veltman's update semantics. We then layer on the modalities, doxastic and deontic operators that are needed to model the various promissory frameworks and their differences in the Promise Update Semantics (PUS). We close the chapter with some simple applications of the framework taken from the informal descriptions in earlier chapters.

4.1 Update Semantics

An update semantics system is a triple, $\langle L, \Sigma, [\,] \rangle$ which consists of a logical language L, a set of information states Σ, and an update function $[\,]$ which assigns to each sentence φ of L an operation $[\varphi]$ on Σ.

Given the state σ and the sentence φ we use $\sigma[\varphi]$ to denote the update of σ with φ. We write $\sigma[\varphi_1]...[\varphi_n]$ to denote the resulting state of updating σ with the sequence of sentences $\varphi_1, ..., \varphi_n$.

Additionally there are two special states, contained in Σ needed for an update system, the minimal state $\mathbf{0}$, and the absurd state $\mathbf{1}$. There are two types of update, successful and unsuccessful. Simply put, an update is successful just in case it does not result in the absurd state, $\mathbf{1}$, and unsuccessful if it does.

Definition 2. (Success)
Let φ be a sentence in the language L, and σ an information state in Σ. The update $\sigma[\varphi]$ is successful *iff* $\sigma[\varphi] \neq \mathbf{1}$.

An important notion in update semantics is that of *acceptance*. Given a sentence φ and a state σ, φ is accepted, written as $\sigma \Vdash \varphi$, in σ if the update of $\sigma[\varphi]$ does not change the state, the agent remains in σ.

Definition 3. (Acceptance)
Let σ be an information state and φ a sentence in the language L. $\sigma \vDash \varphi$ *iff* $\sigma[\varphi] = \sigma$.

With acceptance in hand we can define several notions of *validity* [35, 224]. Here an argument is considered valid if updating the minimal state $\mathbf{0}$ with the premises results in the acceptance of the conclusion.

Definition 4. (Validity)
Let σ be an information state. Consider the argument, in L, where $\psi_1, ..., \psi_n$ are the premises, and φ the conclusion.
$\psi_1, .., \psi_n \vDash \varphi$ *iff* for any state σ, $\sigma[\psi_1]...[\psi_n] \vDash \varphi$.

To move to promissory update semantics we need to define the promissory language, states and updates. In most update semantics the base language is a standard propositional language, for promises we use an enriched language to start.

4.2 Basic Language

We start with a basic propositional language; enriching it with the agent specific modalities, before moving to the epistemic, doxastic and deontic operations.

Let P be a set of finitely many atomic propositions, P. P should be considered as the set of logically possible *actions* that agents can undertake, and therefore promise about.

Our basic language is L_0^P, a propositional language enriched with agent specific modalities. We start with a standard propositional language consisting of a unary operator $\neg$, a binary operator $\wedge$ and two parantheses) and (.

$$\varphi ::= p \,|\, \neg\varphi \,|\, (\varphi \wedge \psi)$$

We then define additional operators in the usual manner:

$$\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$$

$$\varphi \rightarrow \psi := \neg\varphi \vee \psi$$

We will extend this basic system to be *multi-agent*, and incorporate *actionability* and *preference* modalities.

4.2.1 Minimal Promissory State

A *state* is more than just a possible world. The deontic update semantics of van der Torre and Tan introduced the notion of a possible world model representing the state of an update semantics system. Our system will follow suit. Promissory Update Semantics requires frames of states, or *state frames* rather than states simpliciter.

4.2.2 Multi-agency

Let I be a set of finitely many agents. We associate a subset of W to each agent W_i, representing the epistemic state of each agent. In this way we extend the single-agent update semantics to a multi-agent setting.

Each world in an agent's epistemic space, $w \in W_i$, is considered equally epistemically possible to the agent i. If a world is not in W_i then the world is known, by i, to be impossible.

4.2.3 Agent Modalities

We remind the reader how a modal operator works in general. We will have three such operators defined below, for the abilities, preferences and future actions of agents.

A binary relation, minimally reflexive, R is added along with a unary operator $\Diamond$ such that

$$w \models \varphi \textit{ iff } \text{there is some } w' \in W \text{ such that } wRw' \text{ and } w' \models \varphi.$$

We define the dual operator $\Box$ in the usual manner:

$$\Box\varphi := \neg\Diamond\neg\varphi$$

which will have the semantics

$$w \models \varphi \textit{ iff } \text{for all } w' \in W \; wRw' \text{ and } w' \models \varphi.$$

The following modalities are added to L_0^P, enriching its expressibility.

Abilities

We relate propositions P to agents I via the addition of alethic ability relations, R_i^A for each agent. The relation represents the possible actions of the indexed agent. That is, what actions each agent can do. The relations R_i^A are defined as both *reflexive* and *transitive*.

We will use C_i instead of $\Diamond$.

$C_i\varphi$ is read as "agent i *can* (has the ability to) do φ".[1]

[1]The full formal reading of this is "agent i has the ability to make φ true", but the given reading lends clarity and the difference does not cause trouble in the system. We will adopt this slightly sloppier approach for the remainder of the modalities, in the spirit of clarity.

Thus if $C_i(\varphi)$ is true in some world w in σ, then there is some $v \in W$ such that $wR_i^A v$ and φ is true at v. For $C_i(\varphi)$ to be true in some promise state frame σ then all worlds in the relevant W are R_i^A connected to some world which makes φ true.

$C_i(\varphi)$ is true if, $\forall w \in W_i : \exists v \in W$ such that $wR_i^A v$ and $v \models \varphi$.

Preferences

We introduce a simple preference modality for each agent. The relation is a ordering of the agent's, i, dislike for one world over another, R_i^P. The relation is both reflexive and transitive.

We will use I_i which is neither a $\Box$ or $\Diamond$ operator. $I_i\varphi$ is read as "agent i is *indifferent* to φ taking place". If $I_i(\varphi)$ is true at some world w, then for all worlds $v \in W$ such that $wR_i^P v$, $v \models \varphi$.

$I_i(\varphi)$ is true if, $\forall w \in W : \exists v$ such that $wR_i^P v$, or $wR_i^P v \wedge vR_i^P w$ and $v \models \varphi$.

Actions

Here we introduce a third agent specific modality for what an agent *will* do. The binary relation R_i^W is a *tree*. That is, R_i^W is a partial order of W_i and for any $w \in W_i, \{v \in W_i | vR_i^W w\}$ is well-ordered.

(W_i, R_i^W) is a partially ordered set, such that for each $w \in W_i$, R_i^W well-orders the set $\{\}$.

It is used to represent the future actions that an agent will do.[2] R_i^W should be thought of as an *earlier than* relation. Informally, for any two points in the past, they are either equal, or one is earlier than the other.

This will be used when we deal with the doxastic changes of the promisee. We note here that the resulting belief from a promise is not φ but rather the *belief* that the promisor 'will see to it that φ'. That is, the change should be modelled doxastically not epistemically.

We will use G_i, which is neither a $\Box$ nor a $\Diamond$ operator. $G_i(\varphi)$ is read as "agent i will do φ". So worlds that are R_i^W related to a world w, are worlds in which agent i has done φ.

$G_i(\varphi)$ is true if there is some point on each branch of R_i^W where φ is true.

4.3 Extended Language and Update Operators

The basic language L_0^P is extended to the language L_1^P by the addition of doxastic and deontic operators. They are, with rough accompanying definitions

[2] We note explicitly that this is much too simple approach to modelling time. However, the task at hand does not require an explicit temporal modelling. The puzzles that will be discussed in Chapter five are done no harm by this simplistic modelling of an agents temporal actions.

To be technically thorough the model should contain world, time pairs $\langle w, t \rangle$. However, as the model is quite complex as it stands, we opt for this innocent gloss as it adds clarity for the reader.

which will be filled out below:

- $B_i^W(\varphi)$ – "i has *weak* belief that φ"
- $B_i^S(\varphi)$ – "i has *strong* belief that φ"
- $O_{ij}^n(\varphi)$ – "i is obliged to see to it that φ is true, for j under normative force n" where $n \in \mathcal{N}$, the set of normative force indicators.

Definition 5. (Promissory Language)
A string of symbols φ is a sentence in L_1^P *iff* either φ is a sentence in L_0^P or there is a sentence ψ of L_0^P such that $\varphi = B_i^W(\psi)$ or $B_i^S(\psi)$ or $O_{ij}^n(\psi)$.

4.4 Promissory State Frame

For Promissory update semantics we have added a set of agents I and the alethic, preference and ability relations, $R_i^A, R_i^P R_i^W$, we complete the definition of a state with the deontic ideality relations $\leq$, as well as doxastic plausibility relations $\preceq$.

Definition 6. (Promissory State Frame)
A promissory state frame σ is an ideality relation and a plausibility relation represented by a possible world model.

$$\langle W, I, \mathcal{N}, \{W_i\}_{i\in I}, \{R_i^A\}_{i\in I}, \{R_i^P\}_{i\in I}, \{R_i^W\}_{i\in I}, \{\leq_{ij}^n\}_{ij\in I, n\in\mathcal{N}}, \{\preceq_i\}_{i\in I}, V\rangle$$

We do not use the accompanying notion of truth that comes with this possible world model.

Definition 7. (Special States)[3]

> The minimal state, with respect to agent j, is the state in which j considers all worlds, in W, equally epistemically possible, and there are no normative constraints for j, all worlds are equally ideal from j's point of view.
>
> **0**, *the minimal state*, when agent j is in the state,
> $\langle W, I, \mathcal{N}, \{W_i\}_{i\in I\setminus j}, W, \{R_i^A\}_{i\in I}, \{R_i^P\}_{i\in I},$
> $\{R_i^W\}_{i\in I}, \{\leq_{ik}^n\}_{i,k\in I\setminus j, n\in\mathcal{N}}, \{\leq_{jl}^n\}_{l\in I, n\in\mathcal{N}} = W \times W, \{\preceq_i\}_{i\in I}, V\rangle$.
>
> The absurd state, with respect to agent j, is the state in which j considers no world, in W, epistemically possible, and all worlds are no longer deontically linked to each other, from j's normative point of view.
>
> **1**, *the absurd state*, when agent j is in the state
>
> $$\langle W, I, \mathcal{N}, \{W_i\}_{i\in I\setminus j}, \emptyset, \{R_i^A\}_{i\in I}, \{R_i^P\}_{i\in I}, \{R_i^W\}_{i\in I}, \{\leq_{ij}^n\}_{ij\in I, n\in\mathcal{N}}, \{\preceq_i\}_{i\in I}, V\rangle$$

We define the *total* versions of these special states in terms of all agents, in I, being in the minimal (absurd) state.

[3] We represent the special states *epistemically* as, for the purposes at hand, we need not differentiate between epistemic and deontic absurdity (minimality). Alternately we could define an additional *deontic absurd state*, and change the update operations accordingly.

4.5 Promissory Language

The doxastic and deontic states of agents are described by the operators in the extended language L_1^P.

4.5.1 Doxastic Operators

To model the doxastic state of the agents we add a binary plausibility relation $\preceq_i$ for each agent to the promissory state. This relation is reflexive and transitive on W_i.

The use of plausibility orders on worlds is borrowed directly from the DEL family of logics. It represents a ranking of worlds based on their relative plausibility to the agent. The relation is restricted to W_i as agents do not consider epistemically impossible worlds as plausible.

If $w \preceq_i v$ we say "agent i finds world w less plausible than world v".

We define two doxastic operators in L_1^P, indexed by the agent, which rely on the plausibility relation. B_i^W for *weak belief* and B_i^S for *strong belief*.

4.5.2 Deontic Operators

In order to model (prescriptive) obligations we introduce a series of ideality relations $\leq_{ij}^n$, and accompanying operators. First let us speak generally about the operators and relation.

Ideality relations are binary and on a set of possible worlds, i.e. they are a subset of $W \times W$. If $w_1 \leq_{ij}^n w_2$, we say that w_1 is deontically preferred, or ideal, to w_2, in the specified ideality relation.

First we must add the set of normative forces, $\mathcal{N}$. We borrow from Yamada the idea of indexing deontic operators, and relations, by $I \times I$, the Cartesian product of the set of agents with itself. [37, 301] In this way we have separate normative space for each coupling of agents. This is needed as otherwise if Tom promised Sally p, and Geri promised Monica $\neg p$ we would have a normative clash, and reach an absurd state. Thus we have $O_{ts}(p)$ and $O_{gm}(\neg p)$ for the two obligations, where the operators are paired with similarly indexed relations.

We extend this approach by additionally indexing by the set of normative force indicators $\mathcal{N}$.[4]

Just as Tom's promise and Geri's promise should not clash, consider the following.

> *Sally promises Geri that she won't take the next client at work. Tom is the boss and orders Sally to take the next client.*

[4] As Yamada's system in based on a command logic, his modelling of promising in [38] introduces a third agent to the index, such that "it is obligatory upon i with respect to j in the name of k to see to it that". This is motivated by a story of a boss and secretary, with a promise to a client that will be fulfilled by the secretary. Such external factors are not needed, however. The obligation is for i to 'see to it that φ', and there is no specific method involved. Yamada's extension came out of a need for the promisor to have the right sort of authority relationship to enter in an obligation, this is only the case for obligations grounded in *authority*.

While Sally is not in an enviable position, it is clear that different sorts of obligations can clash without leading an agent to an absurd state. Being obligated to p and $\neg p$ can naturally occur under different types of obligations, without being deontically absurd.[5]

We have mentioned various normative forces in passing, thus far, let us enumerate the normative force indicators here.

$\mathcal{A}$ - Authority

κ - Agent of Normative Authority (Deity)

$\mathcal{H}$ - Honour

$\mathcal{L}$ - Legal

$\mathcal{C}$ - Convention

$\mathcal{E}$ - Expectation

$\mathcal{Y}$ - Hybrid

To this list we add one more indicator, for convenience purposes:

$\mathcal{P}$ - Promise

We use this indicator for clarity when the discussion need not differentiate between promissory frameworks.

Similarly we use the separate normative force for the proposed hybrid account. $\mathcal{Y}$ represents the normative force of a convention grounded in expectations, the core of the hybrid account is that we cannot separate the two parts, when dealing with promises, so we leave them as a new tied element, with no technical detriment. Thus the deontic operators are indexed by $(I \times I) \times \mathcal{N}$.

For example we read $O_{ij}^{\mathcal{A}}(\varphi)$ as "agent i is obligated to see to it that φ for j under the normative force $\mathcal{A}$ (Authority)". Note that the order in which the agents are indexed is important. The first agent is the obligator, and the second the obligatee.

If Sally promises Tom that she will p, we have $O_{st}^{\mathcal{P}}$. However if Tom orders Sally to p we get $O_{st}^{\mathcal{A}}$. The commander is the obligatee as discussed in section 3.3.

The approach to modelling promissorry obligations is two-fold.

1. In the initial state all worlds are connected by all deontic relations $\leq_{ij}^{n}$, i.e. for all $w_1, w_2 \in W$, $w_1 \leq_{ij}^{n} w_2$ and $w_2 \leq_{ij}^{n} w_1$, where $n \in \mathcal{N}, i, j \in I$.
2. The addition of obligations cuts links between worlds from the corresponding ideality relation. [33, 4]

The update of the ideality relation by $O_{ij}^{\mathcal{A}}(\varphi)$ "i ought to see to it that φ for j under the normative force of *authority*".

[5] While the present system can model such clashes, it is beyond the task at hand to model how Sally would deal with her conundrum.

4.6 Promissory Updates

Recall that a promise state frame σ is:

$$\langle W, I, \mathcal{N}, \{W_i\}_{i\in I}, \{R_i^A\}_{i\in I}, \{R_i^P\}_{i\in I}, \{R_i^W\}_{i\in I}, \{\leq_{ij}^n\}_{ij\in I, n\in\mathcal{N}}, \{\preceq_i\}_{i\in I}, V\rangle$$

We define an agent substate frame as the set of elements in the promise state frame which are primarily indexed by the agent, and their dependencies. Thus for agent i we have:

$$\langle W, W_i, R_i^A, R_i^P, R_i^W, \{\leq_{ij}^n\}_{j\in I, n\in\mathcal{N}}, \preceq_i, V\rangle$$

In the following update definitions we will use this substate definition, as the updates are all specific to a particular agent.

If the update occurs to more than one agent we will adopt the convention of indexing the set of agents, e.g. $\sigma[\varphi]_{ij}$ for the update of σ by agents i and j learning φ.

4.6.1 Precondition Checks

As we know, one of the things that differentiate promises from other obligation-creating speech acts is that certain conditions must be met before a promise can be made, we need to be in obligation-creating-circumstances. What these circumstances are varies between promissory frameworks. There are two conditions that are included in most frameworks.

i. *(Can)*: The promisor has the ability to perform the promised act.

ii. *(Want)*: The promisee prefers, or is neutral towards, the performance of the promised act.

Normally updates result in a change in the state of the agent. There are also operators that function as a *test*. They check whether a piece of information is consistent with ones knowledge. Veltman introduces this with the *might* operator.

Definition 8. (*might* Operator)

$$\sigma[\mathit{might}\,\varphi] = \sigma \text{ if } \sigma[\varphi] \neq 1$$
$$\sigma[\mathit{might}\,\varphi] = \mathbf{1} \text{ if } \sigma[\varphi] = 1$$

If updating with φ takes the agent to an absurd state, then φ is *inconsistent* with the agent's current knowledge, thus an update with $\mathit{might}\,\varphi$ will also fail. If updating with φ takes us to any state which is not absurd, then there is consistency. As *might* is a test operator, a successful update with it leaves us in the original state, though now aware that φ is epistemically possible.

We will use the *dual* of *might*, $\Box$.

Definition 9. ($\Box$ operator)

$\sigma[\Box\,\varphi] = \sigma$ *iff* $\sigma \models \varphi$,
otherwise $\sigma[\Box\,\varphi] = \mathbf{1}$

Our two preconditions, *can* and *want*, can be used as tests. This is because we have defined both preferences and abilities as *fixed*. The relations R_i^A and R_i^P are assumed at the start of the model, and we do not have any operation which changes these relations. We are able to view them as 'epistemically fixed', and thus they are candidates for test operators.

Let i be the promisor and j be the promisee. *Can* corresponds to $C_i(\varphi)$, while *Want* corresponds to $I_j(\varphi)$. We could construct specific test operators, along the lines of $\Box$ for each case. Note that both target sentences are in the language L_0^P, this means that we can view both checks as specific uses of $\Box$, bringing along other uses from Veltman's update semantics as well.

Definition 10. (Preconditions)
Let i be the promisor and j be the promisee.

(Can): $\sigma[\Box C_i(\varphi)]_j = \sigma$ if $\forall w \in W_j, \exists v \in W$ such that $wR_i^A v$ and $v \models \varphi$, or $\mathbf{1}$ otherwise.
(Want): $\sigma[\Box I_j(\varphi)]_i = \sigma$ if $\forall w \in W_i : w \models \varphi, \forall v \in W$ such that $wR_j^P v$ and $v \models \varphi$, or $\mathbf{1}$ otherwise.

Note that the tests for *can* and *want* are only needed by the promisee and promisor respectively in order to test for obligation-creating-circumstances.

4.6.2 Epistemic Updates

An update is epistemic if it affects the agent's knowledge of the world, i.e. if it changes W_i. Epistemic updates are operator free, i.e. they are in the static language L_0^P.

When an agent learns a piece of (factual) information all worlds that do not contain the information are removed from their epistemic possibilities. It is an *eliminative* approach to epistemic update; as the agents learn the set of possible worlds which could be the case shrinks. If agent i had total knowledge then W_i would be the a singleton containing just the actual world. For the totally ignorant agent j the epistemic space would contain all worlds, $W_j = W$.

Definition 11. (Epistemic Update)
Let σ be a promise state, φ a sentence in L_0^P and $i \in I$. Then $\sigma[\varphi]_i$ is defined as follows.

If $W_i' = \{w \in W_i | w \models \varphi\} \neq \emptyset$,
then $\sigma[\varphi]_i = \langle W, W_i', R_i^A, R_i^P, R_i^W, \{\leq_{ij}^n\}_{j \in I, n \in \mathcal{N}}, \preceq_i, V\rangle$.
and $\sigma[\varphi]_i = \mathbf{1}$ otherwise.

Note that if the agent learns new information that does not fit with their current knowledge they enter the absurd state. We can see why this is the case by the following we update the minimal state with φ, then $\neg\varphi$.

$$\mathbf{0}[\varphi]_i[\neg\varphi]_i$$

The first update will remove all $\neg\varphi$-worlds from W_i, as they will not be members of W_i'. The second update would remove all φ-worlds from W_i, i.e. $W_i = \emptyset$. The agent is epistemically absurd.

4.6.3 Doxastic Updates

An update is *doxastic* if it affects an agent's beliefs about the world, i.e. if $\preceq_i$ is changed. Doxastic update are operator-free, but not modal-free. That is, we allow any sentence of L_0^P to be doxastically updated, but not sentences in L_1^P. It should be clear that $\preceq_i$ is limited to worlds in W_i, as we only hold beliefs over worlds which are considered as epistemically possible.
For Doxastic updates we employ DEL-style notions of *update* and *upgrade*. [31, 13] These correspond to our notions of weak beliefs B_i^W, such as those that result from assertions, and strong beliefs B_i^S, the expectationalist's promissory beliefs. Stronger belief revision policies are used for sources which are more reliable, or trustworthy to the agent [31, 13], [3, 11].

Borrowing from the DEL playbook we differentiate between different types of information flow. We use epistemic updates for updates that are first-hand learnable by the agent. That is, if an agent observes an act taking place then they *know* it to be true. No matter how reliable they find another agent, a report of an action by another agent is strictly doxastic, not epistemic.

For example, if agent i promises to "**pick you up at the airport Tuesday**", and then subsequently does not do so. You know first hand that it did not happen, so the update would be

$$\sigma[\neg(pickyouup)_i]_j$$

Contrast this with you telling Tom that i did not pick you up, here Tom can only incur a belief that this was the case, thus

$$\sigma[B_i^S(pickyouup)_i]_t$$

or if Tom doesn't trust your reporting,

$$\sigma[B_i^W(pickyouup)_i]_t$$

So we will use epistemic updates for first-hand and factive updates, while the doxastic operators will be used active updates and when the agent must rely on the reporting of others.

We first add a *coherency constraint* to the plausibility orders.

Definition 12. (Coherency Constraint)
A plausibility ordering is said to be *coherent* if there is some world which is at least as good as all others. [35, 232]
Let $\preceq_1$ be a plausibility ordering on W_i.

w is a *normal world iff* $w \in W_i$ and $w \preceq_i v$ for every $v \in W$.

$\mathbf{n}_{W_i}$ is the set of all *normal worlds* in W_i.

$\preceq_i$ is *coherent iff* $\mathbf{n}_{W_i} \neq \emptyset$.

We use the operator B_i^S, in L_1^P to represent strong belief change. $B_i^S(\varphi)$ is an update which replaces the current ordering relation with one where all the φ-worlds become better (higher) than all the $\neg\varphi$-worlds, leaving the rest of the ordering intact.

Definition 13. (Radical Upgrade)
Let φ be sentences in L_0^P.

If $\sigma[B_i^S(\varphi)]_i$ is *coherent* then,
$\sigma[B_i^S(\varphi)]_i = \sigma'$, where σ' is the same as σ except that the relation
$\preceq_i$ is replaced by $\preceq_i'$ where,
$w_2 \preceq_i' w_1$ *iff* $w_2 \preceq_i w_1$ and $w_1 \models \varphi$, or $w_2 \models \neg\varphi$,
otherwise $\sigma[B_i^S(\varphi)]_i = \mathbf{1}$.

We also have the notion of *weak belief*, B_i^W in L_1^P. $B_i^W(\varphi)$ is an update which puts the *best* φ-worlds at the top of the ordering of W_i, leaving the rest of the ordering the same.

First we define *best*, as the set of maximally plausible worlds on a given plausibility ordering, $\preceq$:

$$best(W) := \{w \in W \,|\neg\exists v \in W \text{ such that } w \preceq v\}$$

Definition 14. (Conservative Upgrade)
Let φ be sentences in L_0^P.

If $\sigma[B_i^S(\varphi)]_i$ is *coherent* then,
$\sigma[B_i^S(\varphi)]_i = \sigma'$, where σ' is the same as σ except that the relation
$\preceq_i$ is replaced by $\preceq_i'$ where,
$w_2 \preceq_i' w_1$ *iff* both $w_2, w_1 \in best(W_i)$ and $w_1 \models \varphi$, or
not both $w_2, w_1 \in best(W_1)$ and $w_2 \preceq_i w_1$,
otherwise $\sigma[B_i^S(\varphi)]_i = \mathbf{1}$.

Given our general notion of validity, and the coherency constraint it is easy to see that B_i^S entails B_i^W. That is, if an agent updates from a more trustworthy source then the update will contain the result of a more wary update.[6]

[6]Both of the doxastic updates operate on the *same* plausibility ordering. This is because we do not sequester our beliefs based on fallibility of source; rather the extent to which our beliefs, *in totalia*, change is a result of the trustworthiness of the source.

4.6.4 Deontic Updates

An update is *deontic* if it affects an agents normative state, i.e. if $\leq^n_{ij}$ is changed. Deontic updates are operator-free and also modal-free. That is, we allow only sentences of L^P_0 which do not contain modal operators to be in the range of the deontic operators. Recall that the modalities of L^P_0 are of an agent's abilities to perform actions, and their preferences about worlds. These are simply not the correct type for an agent to form obligations about, thus we restrict the range of the operators.' Our deontic operators $O^n_{ij}(\varphi)$ are read as "φ ought to be done by i for j under the normative force of n". The deontic update is done by deleting all ordered pairs in the deontic relation (R^n_{ij}) that connect φ-worlds and $\neg\varphi$-worlds: $w_1 \leq^n_{ij} w_2$ where $w_1 \models \neg\varphi$ and $w_2 \models \varphi$.

Definition 15. *Deontic Updates*
Let σ be a promise state frame, ψ a modal-free sentences of L^P_0, $i, j \in I$, $n \in \mathcal{N}$, and φ a modal free sentence of L^P_0.[7]

> $\sigma[O^n_{ij}(\varphi)]_i = \sigma'$, where σ' is the same as σ except that the relation $\leq^n_{ij}$ is replaced by $\leq^{n\prime}_{ij}$ where,
> $w_2 \leq^{n\prime}_{ij} w_1$ *iff* $w_2 \leq^n_{ij} w_1$ and $w_1 \models \varphi$, or $w_2 \models \neg\varphi$,

4.6.5 Promise Update Semantics (PUS)

One of the advantages of the PUS approach is that we have clear delineation between, and within, agents. The epistemic, or factive, state of an agent i is represented by W_i. Their doxastic state by $\preceq_i$, ranging over W_i. While their deontic state is represented by a series of relations $\leq^n_{ij}$, where i is in the first subscript, ranging over W.

In this way we can clearly show what an agent knows, believes, and the sort of obligation they owe to others. With the system defined, we move to simple applications of PUS to phenomena which we have already encountered.

4.7 What We Can Show

One of the first distinctions made above was that of the difference between assertions and promises. We noted that an assertion does not incur *obligation* in the way a promise does, e.g. through *dependant belief* along an expectationalist reading. That is, a belief that i might perform φ is decidedly different from the larger *reliant* belief that comes with obligation-creating-circumstances. We can model these two phenomena by the use of our two types of doxastic update.

> i "*asserts* φ" has the doxastic result of $\sigma[B^W_j(\varphi)]_j$
>
> i "*promises* φ" has the doxastic result of $\sigma[B^S_j(\varphi)]_j$

[7] We note that we do not require *coherence* in the deontic orderings as we allow normative clashes without the agent proceeding to the absurd state.

An assertion still has doxastic effects, but only in the most plausible worlds. The doxastic change of a promise must involve a more radical upgrade for any account of promising that involves the promisee as dependant on the performance of φ.

We must note that this is not the correct treatment of the doxastic change that comes from receiving a promise. This is because it is missing the *temporal* aspect of promising. When I "**promise to pick you up at the airport**" you do not believe that 'I pick you up at the airport' rather that I '*will* pick you up at the airport'. Thus,

$$\sigma[B_j^S(G_i(\textbf{pick you up at the airport}))]_j$$

The difference between an assertion and a promise was important in the discussion of Scanlon's critique of conventional accounts of promising. We can model the story of the two warriors at the river. Recall that Scanlon's claim is that the non-verbal exchange has all the elements of a promissory convention, thus they engage in simultaneous promises:

$$\sigma[Prom_{ij}^f(\varphi)]_{ij}[Prom_{ji}^f(\varphi)]_{ji}$$

While it is clear that the beliefs of the two warriors have been affected, it is not immediately clear that obligations have been formed, and thus are available to be broken. With the information given, it is not clear that the following is not an appropriate modelling of the puzzle.

$$\sigma[B_j^S(G_i(\varphi))]_j[B_i^S(G_j(\varphi))]_i$$

In our discussion of *vows*, *oaths*, and *pledges* we noted the distinction between the secular and non-secular versions of the phenomena. In order to model the latter we must ensure that we have an added agent in I to represent the normative authority.[8] We will use κ as such an agent.[9]

"*i vows* to φ" $-\sigma[O_{i\kappa}^{\kappa}(\varphi)]_i$

"*i swears an oath* to φ" $-\sigma[O_{ij}^{\kappa}(\varphi)]_i$[10]

"*i pledges* to φ" $-\sigma[O_{ii}^{\mathcal{H}}(\varphi)]_i$

$\text{oath}_{sec} \subset \text{vow}_{sec} \subset \text{pledge}$

oath_{sec} are generally public $-\sigma[O_{ii}^{\mathcal{H}}(\varphi)]_i[B_i^W(G_i(\varphi))]_I$

[8] Though it is doubtful that we will need to use the resulting updates and operators, unless modelling the *ten commandments*, we leave them to be generated out of whimsy and convenience.

[9] To represent 'kami', the Japanese word for spirits, natural forces, or essence in the Shinto faith.

[10] Where j's range includes κ, recall *vow*$\subseteq$ *oath*.

We can even model more complex cases using the normative force indicators. The secular marriage was described as being of two parts, the legal institution and the vows themselves. Thus,

$$\sigma[O^{\mathcal{L}}_{ij}(\varphi)]_{ij}[O^{\mathcal{P}}_{ij}(\phi)]_{ij}$$

Here we have the obligation that stems from the institution, $\mathcal{L}$, as well as the obligation that stems from the promise(s). Recall we use $\mathcal{P}$ as a generic placeholder for promissory obligation.

Commands are similar speech acts to promises, we can see the differences between them. A command comes from an authority and obligates the commandee. It does not depend on any doxastic change, and the only precondition is that the authority relationship exists. We have not yet developed this precondition in PUS. A possible approach would be to add an authority ranking of all $i \in I$, for each agent. An agent j could be said to 'have authority over i' if they were greater than (or equal to, for those who believe in self-authority generated obligations) i in i's ranking. Let $auth(i, j)$ be such a test operator.[11]

Obligations from promises, on the other hand, are grounded in *expectation*, *convention*, or a combination of the two. They also require the preconditions of *can* and *want* to be in obligation-creating-circumstances.

"i *commands* j to φ" $-\ \sigma[auth(i,j)]_I[O^{\mathcal{A}}_{ji}(\varphi)]_j$

"i *promises* j to φ" $-\ \sigma[\Box I_j(\varphi)]_i[\Box C_i(\varphi)]_j[O^{\mathcal{P}}_{ij}(\varphi)]_i[B^S(G_i(\varphi))]_j$

4.7.1 Modelling Promise frameworks

A promise is a speech act that has both illocutionary, and perlocutionary effects; these effects are deontic and doxastic. Treating these effects as successive updates allows us to model promises within the update semantics, as seen in the previous section. So we arrive at a new slogan:

"A promise is a chain of updates."

We can now model promises in the different frameworks. We introduce shorthand for the various accounts of the form $Prom^f_{ij}(\varphi)$ for "i promises j to see to it that φ", where f is some promise framework.

Expectionalism

The heart of the expectationalist accounts of promising is the grounding of the obligation in the *reliance on the belief* that φ. We can define what a promise is, according to the expectationalist. Recall B^S_i is stronger belief update than B^W_i, and that for *coherent* plausibility orderings it entails B^W_i. We use B^S_i to

[11] We note that we have the distinct advantage of avoiding Ross' paradox as $C_i(\varphi)$ and $C_i(\varphi \wedge \psi)$ are distinct. While this is a mere sketch of handling imperatives, and other familiar paradoxes may still rear their heads, the path starts with fruitful returns.

model reliant belief as it is more difficult to give up φ once an agent updates with $B_i^S(\varphi)$, due to the total reordering of φ-worlds over $\neg\varphi$-worlds.[12]

"A promise is an obligation from a belief."

$$\sigma[Prom_{ij}^{e}(\varphi)]_{ij} := \sigma[\Box I_j(\varphi)]_i[\Box C_i(\varphi)]_j[O_{ij}^{\mathcal{E}}(\varphi)]_i[B_j^S(G_i(\varphi))]_j$$

Conventionalism

The conventional frameworks do not pivot on the beliefs of the promisee, though they do require uptake. The normative force stems purely from the *convention* of promising.

"A promise is a conventional obligation."

$$\sigma[Prom_{ij}^{c}(\varphi)]_{ij} := \sigma[\Box I_j(\varphi)]_i[\Box C_i(\varphi)]_j[O_{ij}^{\mathcal{C}}(\varphi)]_i[B_j^W(G_i(\varphi))]_j$$

We note here that j ranges over $I\backslash\{i\}$ as the conventional rules require a distinct promisor and promisee.[13] We include the weak doxastic update to ensure that uptake occurs for the promisee.

Hybrid Accounts

The hybrid accounts share the idea that the obligations of promises are grounded in both the convention and the expectations that results.

The *conventional expectations* hybrid account splits the promissory family of speech acts into the narrow type of *promise* and the wider category of *pledges*. The latter was covered in our discussion of other promissory acts, in section 4.7. On this account the obligation is not split between expectations and convention, i.e. they are equal parts of a promissory obligation.

"A promise is an obligation and a belief."

$$\sigma[Prom_{ij}^{y}(\varphi)]_{ij} := \sigma[\Box I_j(\varphi)]_i[\Box C_i(\varphi)]_j[O_{ij}^{\mathcal{Y}}(\varphi)]_i[B_j^S(G_i(\varphi))]_j$$

[12] Radical Upgrade is often cited, in the DEL literature as the doxastic operation for fallible, but trustworthy sources. [31], [3].

[13] While we could insert the restrictions directly on the operators, as alluded to in the discussion in section 4.7, it is methodologically appealing to leave the Update Semantics uncluttered and restrict instead on the application.

Chapter 5

Promising Puzzles - Applying PUS

Now that we have developed the logical framework, and defined promising under the various frameworks, we can examine common puzzles of promising to show how the different philosophical approaches differ under these notions of promising.

In each case we will present the puzzle, provide an initial modelling using $\mathcal{P}$ as a place-holder for the normative force indicator, and then we will discuss how the accounts of promising deal with the puzzle philosophically, then formally.

5.1 Silent promises

> *It is Christmas morning, the family is about to start opening presents. The father watches from the other room as his daughter opens her present. He catches the wave of disappointment wash across her face: it is not the Lego that she has been pining after. At that moment he, quietly in the other room (or not aloud at all), says "**I promise that next year I will get you a gift that you want.**"*

On the surface this has the makings of a promise. We see that the father has the ability to perform the act, $C_i\varphi$. Similarly we can, safely, assume that the daughter would like to receive a preferred gift, $I_j\varphi$. The words 'I promise' are uttered and an obligation results for the speaker. The *silent promise* is called as such because of the situation lacks uptake. Let us try and model the situation in our technical framework.

$$\sigma[O^{\mathcal{P}}_{ij}(\varphi)]_i$$

The silent promise has commonly been used as a problem case for expecatationalist accounts. For expectationalists the normative force of the promise is derived from the dependance of the promisee. It is clear here that the daughter

has no knowledge that the promise exists, so there are no perlocutionary effects. Recall,

$$\sigma[Prom^{e}_{ij}(\varphi)]_i = \sigma[O^{\mathcal{E}}_{ij}]_i[B^{S}_{j}(G_i(\varphi))]_j$$

As the silent promise is lacking the latter half, the belief change, and the deontic fact of the act is limited to the promisor, i, it seems that such objections are not out of place.

Conventional accounts have similar problems with silent promises. On Hume's strict account we see that criteria *(5) ... there must be mutual recognition by both parties that a promise has been made* is directly violated. Given that the utterance is silent, or at least unheard by anyone else, we may run into trouble on criteria *(2) that we indicate we are using the institution when we say '**I promise**'.* The key being the lack of uptake of the indication.

$$\sigma[Prom^{c}_{ij}]_I = \sigma[O^{\mathcal{C}}_{ij}]_I$$

We see that the obligation in question being limited to i, remains a problem in this case.

It should be unsurprising that we find similar problem with the proposed hybrid framework. However, we do not want to say that the father has failed to be in obligation-creating-circumstances, and thus unobliged when uttering a silent promise.

We can explain the silent promise if we widen our search into other promissory acts. That is, we view the utterance of, 'I promise', as an example of a *stretch*, and not a clear indicator of the normative phenomena taking place. Our initial modelling is an obligation from agent i to agent j, where the deontic fact is only updated in the promisor. In structure this is very close to our characterisation of a *pledge*, $\sigma[O^{\mathcal{H}}_{ii}(\varphi)]_i$. The problems of promising many; unknown uptake, possible refusal etc. are mirrored here. This casts serious doubt that the obligation is between i and j. There is no way the promisee can exercise their *right of performance*. The promisor is likely, in the event of breach, to feel dishonoured, which is in line with our characterisation as a pledge.

One may argue that once the promisee is made aware of the silent promise the obligation, right of performance, and punishment criteria all line up with our standard notion of promise, and thus the silent promise is not a pledge. There are two ways in which the promisee could learn of the silent promise. First, and most likely, the promisor could inform the promisee at a later time that this was the case. This is plausible, and will trigger the necessary doxastic change, as well as the epistemic update of the deontic fact. However, it is more likely that the promise is said to have taken place at this reveal, i.e. this cannot just be just a factive update, as it has perlocutionary effects. So, the promisor has converted a pledge into a promise by inviting reliance, on the part of the promisee, and they also invoke the promising convention by their utterance of 'I promised that...', though it is embedded in a more complex temporal statement.

It could also be the case that a third party, who was also in the other room and overheard the utterance, informs the promisee of the existence of the promise. However this would not seem to be obligation-creating-circumstances,

the promisee is unlikely to have a reliant belief change based on the testimony of a third party. This type of relation would most likely be treated as an assertion, thus it may trigger a change in the promisee's beliefs, but only on the *best* worlds, $\sigma[B_j^W(G_i(\varphi))]_j$.

Despite the use of the phrase 'I promise' within the silent promise, an examination of the illocutionary and perlocutionary effects shows that it is not a promise. The phenomenon can accurately be captured when it is viewed as a pledge, as shown by the formal modelling. This mapping is supported if one considers what changes would occur if the father explicitly pledged, rather than promised, that he will get a better gift for his daughter next time. It seems not much more than the utterance, the normative force seems more in line of the father's personal honour and the availability of punishment and waiving the obligation support such a claim.

5.2 Perceived Promises

> *A recruiter and an applicant are in the midst of the interview process. The recruiter* mentions *the positive experiences of recent hires.* ***"Previous new hires received promotions within their first three years of tenure with our organisation."*** *Sadly, for our hire, a promotion does not come to pass within the three years.*[1]

Opinions on the nature of *perceived promises* are split. They are often called 'implicit promises', and this label sheds some light on the phenomena. D. Rousseau argues that the context of the speech act, a formal interview, construes a promise that the applicant will have similar experiences if hired. That is, the interview is a correct set of obligation-creating-circumstances.

While this may seem troublesome for the conventionalist, need not be so dire. The convention only requires *(2) that we indicate that we are using the institution when we say 'I promise', or some equivalent.* Thus we can fit Rousseau's claim into this framework by stating that the nature of the interview implies claims made be the recruiter to be of the necessary sort. A similar move is available to the expectationalist. In the context of an interview, recruiter claims invite reliant beliefs.

S. Bankins offers the following argument against the notion of implicit promises.

1. There is no commitment to by the recruiter to undertake a future action.
2. It is not under the recruiter's control to do so.
3. It is unclear what the obligation would be, if anything.
4. It is unlikely that the applicant will be reliant on the claim. [4, 5]

As the recruiter is offering only a factive description of past events, there is no φ that can be easily pointed to as content of obligation, let alone whether

[1] Adapted from [24, 527] via [4, 4].

it would be of the form "i will see to it that φ". Even if one could work out what the exact obligation is, it is highly unlikely that the recruiter will be able to exercise such control. In the modern interview process recruiters are often member of the Human Resources (HR) department, or even third-parties. In most organizations the HR department does not have the ability to promote, though they may be able to negatively influence promotions based on performance, complaints, etc. As these facts are known, the interview process cannot be said to have *implicit* obligation-creating-circumstances. Thus we do not have a promise under the conventional account.

$$\frac{[\neg C_i\varphi]_I \qquad \neg O^{\mathcal{C}}_{ij}(\varphi)}{\neg Prom^{c}_{ij}(\varphi)}$$

From the expectationalist corner, it is highly unlikely that the applicant "would, say, purchase a house based upon a belief that he or she will indeed receive a promotion within the first three years of tenure". [4, 5] So, while the applicant may well generate a belief about the promotion, they do not generate a sufficiently reliant one.

$$\frac{[\neg C_i\varphi]_I \qquad [B^W_j(G_i(\varphi))]_j}{\neg Prom^{c}_{ij}(\varphi)}$$

It is clear, that the example fails on the preconditions of promising, but it in interesting that it also fail, in different ways, in other ways as well. It should be clear that this is not a promise in the *hybrid* account, either.

We can show Rousseau's account as:

$$\sigma[\Box C_i\varphi]_I[O^{\mathcal{C}}_{ij}(\varphi)]_{ij}[B^S_j(G_i(\varphi))]_j$$

That is we can, in the event of the purchase of a house, explain the misunderstanding as an update of $[B^S_j(G_i(\varphi))]_j$ instead of $[B^W_j(G_i(\varphi))]_j$.

5.3 Deathbed

> *Your ailing grandmother is in hospital. During one of your visits she calls you aside and asks you to take care of her dear doddering dog, in the case that she passes on. You reply with "**I promise you that I will take care of Toby if anything happens to you.**" A few weeks later, your grandmother passes away.*

Most people would respond that you are now bound to take care of poor little Toby, as you promised precisely that if something happened to her that you would take care of him. That is, you are under the following obligation. Let φ be 'take care of Toby'.

$$O^{P}_{ij}(\varphi)$$

While this sort of promise seems straightforward there is the distinct problem of the lack of a promisee. Most would be compelled to say that this promise is more binding than a standard promise with a, still-living, promisee. Consider the alternative:

> *Your grandmother is getting on in years, and is no longer able to take care of her peppy pooch. She asks you to take over care of Toby. You "**promise to take care of Toby**".*

The operative question here is whether the two promises have the same binding strength, or is one lesser than the other.

Many consider the binding strength of a promise to the recently departed as stronger than one to a still living body. This can be explained by the fact that there can be no petition for absolvement. We observed that if circumstances change that the promisor has a legitimate cause to petition, and expect absolvement. The promisor always has the right to petition, but the expectation of absolvement requires a change in the circumstances that would have made the initial promise not in obligation-creating-circumstances. Given that the promisee is not available, in the first instance, this option is not open, and could explain the intuition that such promises are more binding: *there is no escape route.*

This solution is promise framework agnostic, as it relies on an evaluation of the obligation-creating-circumstances, which are uniform. There are further problems with deathbed promises and the specific frameworks.

The expectationalist has the full normative force of an obligation resting with the expectations generated in the promisee. Without reliant belief a promise is not said to occur in the first place. So we must ask what happens when said belief disappears. Let us consider an alternate story that retains the lack of belief but does not have the above issue of petitioned absolvement. We borrow from Downie.

> *You notice that your next door neighbour's lawn is quite overgrown, being a good Samaritan you go over and **promise to mow her lawn the following day**. Your neighbour is getting on in years and he is senile. That evening you see him on his porch and say you will deal with the lawn in the morning. He responds with a puzzled expression and asks what you are talking about.* [11, 265]

Here we have a *forgotten promise*, rather than a deathbed promise, however the problem for the expectationalist is the same in both cases. There is no reliance of the promisee on the promisor, thus no promise can be said to exist.

We can model this as follows:

$$\sigma[Prom^{e}_{ij}(\varphi)]_{ij}[\neg B^{S}_{j}(G_i(\varphi))]_j$$

That is, we have the following dependancy problem.

$$\frac{[\neg B^{S}_{j}(G_i(\varphi))]_j}{[\neg O^{\mathcal{C}}_{ij}(\varphi)]_j}$$

In both the deathbed and forgotten promises the expectationalists need to tell a further story in order to explain how the obligation does not dissipate when the tied belief does as well. Thus it seems that the expectationalist can have promissory obligations disappear without being explicitly waived.

Now consider the conventional account. Here a promise does not need any form of belief on the part of the promisee, though commonly at least weak belief would be expected.

$$\sigma[Prom^c_{ij}]_{ij} = \sigma[O^{\mathcal{C}}_{ij}]_i$$

An update with $\neg B^S_j(\varphi)$ does not change the deontic state of agent i, thus we do not have a problem with the forgotten promise.

There is a technical wrinkle with the deathbed promise, however. In this case we must ask the overarching question of what to do when an agent j dies. Do we remove them from I? If this is the case, then all relations indexed by j would be vacuous and have to be removed as well. If this tactic was adopted, then the conventional accounts would also have a problem with such promises.

First let us recall the conventional requirements of a promise: *(9) There is always a promisee.* At the time of the promise, this was certainly true, however this seems not enough to clear the deathbed hurdle.

Under Hume's account, promising is a communal institution, and as such part of promise-keeping is for the benefit of the community in which the institution resides. The obligation to keep promises begins from self-interest but becomes moralised through sympathy with the public interest. [11, 260] In this way it seems that the conventional account avoids the trap of the dissolved deathbed promise.

We are still left with the problem of what to do *formally* to your grandmother. One solution would be to place such an agent directly into the *absurd state*. This would prevent the agent from being able to update, while not dissolving the deontic relations involving them.

For the hybrid account we rely on the conventional solution to tide us through the lack of expectation, noting that the two cannot be separated in a hybrid view, so both must fail for the promise to also do so.

5.4 Duress

> *A member of the I.R.A., cornered by the police forced his Protestant hostages to* ***promise on their bible not to give evidence against him to the police****, before ultimately fleeing to parts unknown. The hostages, unsure of what to do consult their minister. He advises them that they had an obligation to keep their promise.* [11, 262]

From this example, a modern version of Smith's highwayman, we can see how the three promissory frameworks differ, as well as identify as to what school of promising the minister attends. These questions are left as an exercise to the reader.

Chapter 6

The Logical Form of Promising

Up until this point we have only discussed simple promises in the form of atomic sentences. We can, of course, promise sentences of more complex structures, though these are not uniformly treated in a propositional manner.

There are two ways in which a connective can affect a promise, within the scope of the promise or outside of it. We start with the simplest, the negation.

6.1 Negations and Promises

It is clear that one can promise a negation, where the negation exists within the scope of the promise operations. What about a *negated promise*, where a promise is directly within the scope a negation?

i. $\sigma[Prom^f_{ij}(\neg\varphi)]_{ij}$

ii. $\sigma[\neg(Prom^f_{ij}(\varphi))]_{ij}$

The former is trivial, it is a promise that "i will see to it that φ is not the case for j". The latter, however, is less straightforward. Recall that promises are obligation-creating utterances. What is it to promise to not create an obligation, or not create a reliant belief that φ. We see that a negated promise cannot be explained theoretically, and that it should be semantically blocked by any system.

6.2 Conjunctive Promises

It is obvious that one can promise two things within the same utterance, how that is modelled is an open question. There are two options for modelling "**I promise φ and ψ.**" The two conjuncts could be treated as separate promises,

i.e. sequential updates. Alternatively they can be treated as a conjunction within the scope of the promise.

i. $\sigma[Prom^f_{ij}(\varphi)]_{ij}[Prom^f_{ij}(\psi)]_{ij}$

ii. $\sigma[Prom^f_{ij}(\varphi \wedge \psi)]_{ij}$

Compare the following:

> *A mother is in discussion with her daughter about various and sundry topics. She says "**I promise that we'll have pizza tonight for dinner, and that you will get a raise in your allowance next month.**"*

> *A mother and daughter are talking about dinner plans. The mother says "**I promise that we will go to FEBO for dinner and then Dairy Queen for desert.**"*

For the first example if the two promised acts are to be taken as one promise, as in (ii), then if the family did not have pizza for dinner the raise in allowance would already been broken, before next month occurred. It is much more plausible that these are two separate promises made within the same utterance. That is option (i) appropriately models the first scenario, as the actions may be presented together but they represent two distinct promises, evaluated separately.

In the second example it is clear that if the family doesn't go to FEBO, or Dairy Queen then the promise is broken. That is, for the promise to be kept both conjuncts must take place. Thus the promise is modelled by alternative (ii).

So we see that there are two types of conjunctive promise, the conjunction of promises and the promise of a conjunction, both fit easily in our system.

6.3 Disjunctive Promises

Promising a disjunction is less clear than the promise of a conjunction. Promises are viewed as either an inducement to perform an act, or an invitation of reliant belief that an act will occur. As a disjunction is true if only one disjunct is true both options seem ill-fit with such semantics.

We do note that the options presented for conjunctive promises do not seem apt for the disjunctive. While we can have a conjunction of promises, the disjunctive analog is untenable.

$\sigma[Prom^y_{ij}(\varphi \vee \psi)]_{ij}$

$\sigma[Prom^y_{ij}(\varphi) \vee Prom^y_{ij}(\psi)]_{ij}$

A promise is an obligation-generating speech act. If we had a disjunction of promises, it would not be clear to what obligation one would be under. Moreover

the promisor would seem to have the option to what obligation they needed to follow through on. So we are left with only the first option.[1]

There are two types of disjunctive promise, the *temporal* and *atemporal*. Atemporal disjunctive promises have no temporal relation between the disjuncts.

> **I promise that we will go to Disneyland, or have lasagne for dinner.**

Here we see the above-mentioned problem of relevance for expectations. Not all such promises are so clear. If the two disjuncts are contextually linked we do not encounter this difficulty.

> **I promise that we will either go to McDonalds or Dairy Queen for dinner.**

The two options are similar, they are about the same event, and do not trigger a strong expectational gap. We can think of these types of promise as a promise about an event, *where we will have dinner* with a list of relevant choices.

The conventionalist story is similar here, trading linked expectations for related acts being induced. Our framework has no way of tracking the relevance of disjuncts, as this would need some form of strong contextualism of sentences. We leave this as a known gap in PUS.

The second type of disjunctive promise is the temporal, that is there is some temporal relationship between the two promised actions. This means that temporal disjunctive promises are not commutative, they are more in line with conditional promises.

In *Promises and Threats with Conditionals and Disjunctions*, van Rooij and Franke discuss both disjunctive and conditional promises. The underlying theory of promise, and threat, is that of promises-as-inducement, where the speaker is attempting to influence the future actions of the hearer. Thus, only the temporal varieties are under discussion.

They show that disjunctive promises are riskier than their conditional counterparts, through propositional equivalence, and thus are viewed as threats, not promises.

Consider the following sentences:

a. If you give me your wallet, I will reward you splendidly.
   $A \to R$ (promise)

b. You will not give me your wallet or I will reward you splendidly.
   $\neg A \vee R$ (threat)

b. is read as a threat because the mention of a possibility raises its salience. The reading of a disjunction equally weighs both disjuncts, the truth of either entails the truth of the whole sentence. However, one can present such promises either as disjunctive or conditional, via the logical equivalence of $\varphi \to \psi$ and $\neg\varphi \vee \psi$.

[1] This is already reflected in PUS, as we do not allow complex updates which contain the dynamic operators, i.e. sentences of L_1^P.

The conditional form mentions the desirable antecedent, φ, while the disjunctive form mentions the undesirable negated form, $\neg\varphi$. Disjunctive temporal promises are often viewed as threats, not promises, because of this difference in mentioning. It is a riskier move for an agent to bring up the undesired, versus desired, decreases the utility of the speaker. The use of temporal disjunctive promises is thus a suboptimal strategic inducement. [34, 14]

A threat is cheaper than a promise, in terms of expectations, and this is why the disjunctive promise is more risky than the conditional threat. Van Rooij and Franke conduct their analysis under the presumption of an expectational reading of both promises and threats, but we see that the results are the same in conventional flavours as well. They view promises and threats as strategic commitments to induce behaviour in the hearer, and they adopt a game theoretic analysis to explore these commitments.

This analysis shows that we can have no disjunctive promises, despite their logical equivalences to sentences in the conditional form.

6.4 Conditional Promises

A common form of promising is that of the *conditional promise*. The promised act is contingent on a certain circumstance being the case, before the promised act needs to take place.

> **If it rains, I promise to pick you up at the airport**

There are two types of conditional promises, based on the antecedent. The above example is of a *situational* conditional promise. The obligation to perform the promised act only arises in a certain set of situations, described by the antecedent of the conditional. The antecedent is *factive*.

The other sort of conditional promise is the *active* conditional. Here the antecedent is the performance of a some act by the promisee. The obligation for the promisor to perform the promised act relies on the performance of some prior act by the promisee, outlined by the antecedent of the conditional.

> **I promise that if you buy me a pint, then I will edit your thesis.**

The active conditional promise differs from the situational as the agents retain full control of both the antecedent and the consequent of the conditional. This allows the promisee another avenue to dismiss the promise, by not performing the initial act the promisee ensures that the promisor is not obligated to perform the promised (consequent) act.

As the antecedent is an action this means we must buttress our requirements for obligation-creating-circumstances. Let our promise be $\varphi \to \psi$, i.e. φ 'you buy me a pint' and ψ 'I edit your thesis'.

Our usual tests are $C_i(\psi)$ and $I_j(\psi)$. In addition we must perform similar tests for the antecedent act: $C_j(\varphi)$ and $I_i(\varphi)$[2] For the situational variety the check would simply be one of metaphysical possibility.

Conditional Scope

In the case of conjuctions and disjunctions we saw that there are different ways to model a promise of those forms, for the conditional we too have two options. Taking our original conditional promise let φ be 'it rains' and ψ 'I will pick you up at the airport'. We have two options for modelling.

i. $\sigma[Prom^f_{ij}(\varphi \to \psi)]_{ij}$

ii. $\sigma[\varphi \to Prom^f_{ij}(\psi)]_{ij}$

G. R. Grice offers the following description of conditional promises which sheds light on our choice.

> It is only when these conditions are satisfied that we can say that G did promise, simpliciter... he is under an obligation if these conditions are fulfilled. And ... from the proposition that G promised it follows that he is under an obligation–for we can say that he promised only if the conditions upon which he promised are fulfilled. [13, 56][3]

Grice, then, is an advocate of the second option. A clear advantage to this reading is that it prevents vacuous promises. If the scope of the promise ranges over both the antecedent and consequent then the conditional is always considered true, and thus the promise in tact, if the antecedent is false: $\neg\varphi$. By restricting the scope Grice avoids this problem.

There is a problem with this option. W. R. Carter notes the above advantage but points out a deeper issue. If we view the scope of the promise as restricted to the consequent then we are left with no obligation being formed at the moment of promising. With both the situational and active conditional varieties, a promise is only made when the antecedent condition is satisfied. The world obligated the promisor at some point in time.

> [I]f I say on Thursday that I promise to take my son to the museum on Saturday, what I must mean is, not that I promise, given certain conditions, to take him, but that if certain conditions are fulfilled,I then promise to take him. [10, 32]

This seems counter-intuitive, surely the obligation is created at the moment of promising, or not at all. Carter offers a possible reading: *nested promises*. A

[2] Arguably $I_i(\varphi)$ could be assumed as the promisor is the utterer and the contrary would be quite obtuse.

[3] via [10, 31].

conditional promise is a promise that if φ is the case then one promises to do ψ.

$$\sigma[Prom^{f}_{ij}(\varphi \rightarrow (Prom^{f}_{ij}(\psi)))]_{ij}$$

This solution covers the gap of non-obligating at the time of utterance only to expose another. Here one obligates at the time of utterance that they *will* obligate once a certain condition is met. That is, the core promise to perform the promised act is still only formed if the antecedent condition is met, at some nebulous time in the future. Here Carter concludes this approach is untenable, surely the promise to 'see to it that ψ' must also be formed at the time of utterance, the nesting trick merely obscures the problem.

We note that there is a way of saving the nesting approach, we could detach promising and obligation. That is we could appeal to a future obligation to rest in the consequent, with the scope of the *promise* remaining across the entire conditional. In this way you obligate, under the promise in the form of (i.) but there is a secondary obligation which satisfies Grice's approach for obligation. Vacuous promises, conditional promises with unfulfilled antecedents, hold, but the obligation to perform the act is left alone in these cases.

Such an approach gets around Carter's problem, while retaining most of Grice's advantage, however it is more of a technical manoeuvre. The real problem at hand is *temporality*.

Promises are about the performance of *future acts*. In order to correctly model the situation we must somehow isolate the antecedent situation and then deal with the obligation to the consequented act. Both Grice and Carter attempt to model the promise within a first-order-like structure. Let us look closer at the bounds of such an approach

Adequacy of Propositional Logic

In a series of articles focusing on human reasoning, [5] [6] [7], the psychologist Sieghard Beller explored conditional promises and threats.[4] The thrust of these articles is that propositional logic is ill-equipped to deal with all the facets of such phenomena. Beller is concerned with conditional *inducements* and how we can appropriately model them.

The conditional is uttered, by the promisor, to bring about a certain action φ by the promisee. This is done by offering to perform some action ψ if the promisee performs P. The conditional promise, then, is a statement declaring that action φ is a sufficient condition for the performance of action ψ.

Most often this is displayed as a simple conditional statement, like our example "If you buy me a pint, then I promise to edit your thesis. ψ is only performed after φ has been induced, and not vice versa. The temporal relation between φ and ψ is unidirectional. Because of this temporal nature of the promise, it can only be broken by the promisor and not the promisee. This is

[4]The kernel of this section grew out of a course paper written in 2008, during my undergraduate degree in philosophy. [19]

despite its predication of an action for both parties. It is only once the promisee has preformed φ that the promisor must perform ψ.

Following Beller, let us focus first on the conditional nature of the utterance, and we will deal with the promissory component in due course. Given any conditional, there are four potential inferences that can be made:

i. *Modus Ponens* (MP): $(\varphi \to \psi), \varphi \Vdash \psi$

ii. *Modus Tollens* (MT): $(\varphi \to \psi), \neg\psi \Vdash \neg\varphi$

iii. *Affirmation of the Consequent* (AC): $(\varphi \to \psi), \psi \Vdash \varphi$

iv. *Denial of the Antecedent* (DA): $(\varphi \to \psi), \neg\varphi \Vdash \neg\psi$

Only the first two of these inferences, MP and MT, are logically valid; the other two are not classically valid. In our example, if you buy me a pint, then I am obligated to help you with your thesis. This is a straightforward instance of MP. If I do not help you with your thesis I have broken my promise.

Beller notes a problem with this modelling. Namely what do we say if you never buy me the pint in the first place. Usually this would mean that I am not going to help you with editing your thesis. If I were to help you with your thesis all along, I could not offer it as an inducement to get you to buy me a pint, promising on pain of. If we accept this reasoning, we are forced to accept an instance of *denial of the antecedent*, an invalid inference pattern. [7, 113]

Beller offers a solution to the problem of DA; interpret conditional promises as *biconditional* statements, thus the lacking reliance between the performance of the two acts is solved.

> **I promise to help you edit your thesis if and only if you buy me a pint.**

If one or the other act is not performed the biconditional is falsified. Note that if both acts are not performed then we are left with a 'true' promise, in the sense that it is emptily waiting for the inducement to take effect. This lines up with the idea of conditional promises (and threats) as inducements to act . [7, 210]

The biconditional treatment may solve the problem of the invalid inference of DA, but it brings along its own problems. Because of the biconditional if you do not buy me a pint, then I *should not* help you by editing your thesis. When presenting the problem of DA, it was because of the manipulative nature of conditional inducements. Helping with the thesis would move me into a similar category as *empty threats*. While such breaches would likely weaken my ability to induce behaviour in you in the future, as with the case of threats, we do not want to classify this the same as breaking the promise to help, when the pint is adequately delivered. [19, 8]

Modelling a conditional promise as a propositional sentence, either as a material conditional or a biconditional, misses the *temporal* relationship between the two acts. For the material conditional, we are not guaranteed that there is any temporal spacing between the antecedent and the consequent; they could

occur at the same time. With the biconditional this problem is exacerbated as the two propositions have no temporal spacing. $\varphi \leftrightarrow \psi$ is equivalent to $\psi \leftrightarrow \varphi$, but this is not the case in the situation at hand. If I help you with editing your thesis *before* you buy me a pint, this does not then obligate you under the promise to buy me a pint, I have helped you *outside of the promise*.

6.5 Adding Temporality

In chapter two we discussed the nature of promises was of *future acts* by the promisor. Up to now we have made the temporal nature of promises implicit in the apparatus. In the following discussions time will be a factor. Our formal semantics is not yet equipped to deal with temporal matters, much like the propositional calculus. We could extend the formal system further to also include temporal relations and operators, more complex than our *will* modality.

Instead let us try an alternate approach, which does not require any technical changes so much as a (slight) conceptual shift. Consider the modelling of the "**I promise I will pick you up at the airport**". Formally we have,

$$\sigma[Prom^f_{ij}(\varphi)]_{ij}$$

We see that before the promise was made we were in state σ, and afterwards the state shown above. If we continue the story, in which you reveal that you are quite taken with trains these days and would prefer not to be picked up, we would model this with a further update $[\neg I_j\varphi]_{ij}$. The full modelling of the story would then be:

$$\sigma[Prom^f_{ij}(\varphi)]_{ij}[\neg I_j\varphi]_{ij}$$

We can think of updates as *events* in the promissory timeline. We have the starting state σ, and then the promise is made. Recall that a promise is a chain of updates, so our timeline looks like this:

$$\sigma[O^{\mathcal{P}}_{ij}(\varphi)]_{ij}[B^S_j(\varphi)]_{ij}[\neg I_j\varphi]_{ij}$$

So we consider each update $[\Phi]$ as a distinct moment in time. One can immediately see a problem with this treatment of time. Up until now we have been bundling the doxastic and deontic effects together, hence our employment of the shorthand $[Prom^f(\varphi)]_{ij}$, however this approach to temporality would see these as separate events. Our response is that due to the nature of the speaker uttering, their deontic effects are slightly ahead of the hearer's doxastic update. This may seem a tenuous appeal, but upon further examination it reduces to a mere quirk. It is uncontested that the deontic update happens slightly ahead of time. One could argue a stronger case that the deontic effect happens during, or at the beginning of, the speech act. The problem with such a claim appears to be one of scope. We are appealing to a loose event-based temporal schema on the one hand, then speaking in milliseconds on the other. Rather than seeing this as a breach, this helps underline the weakness of the notion of temporality. Let us label this concept:

One Step at a Time Principle (OSaaT):
Updates happen at unique moments in time; the history of updates can be viewed as a promissory event time-line.

Under this conception we cannot speak of the (temporal) space between events; whether they be mere milliseconds, or years. For the purposes at hand this is not a handicap, for we are solely concerned with the promissory events which occur in the timeline. This will become clear in our discussion promissory events that surround the *performance* of the promised act.

6.6 Modelling Conditional Promises

In order to model conditional promises we must show the temporal relationship between the antecedent and the consequent.

The deontic update semantics of van der Torre and Tan takes the conditional obligation as primitive. Recall that DUS is a single-agent system that deals with epistemic and deontic updates in a similar fashion as to how we deal with them in PUS, the latter being based on the former.

A deontic state in DUS is a possible worlds frame:

$$\langle W, W^*, \leq, v \rangle$$

Where W is the set of possible worlds, W^* a subset of W representing the agent's epistemic state, $\leq$ a reflexive (but not transitive) binary relation on W, and the valuation function V.

The deontic operators in DUS are not unary as they are in PUS, but *dyadic*. The operators come in two forms $\mathbf{oblige}(\alpha|\beta)$ and $\mathbf{oblige}^*(\alpha|\beta)$, where α and β are sentences in a standard propositional language. They refer to the reduction of the ideality relation on W and W^* respectively. The sentence $\mathbf{oblige}(\alpha|\beta)$ can be read as "α ought to be done if β is done". That is the obligation to do α only exists in worlds in which β is the case. [33, 7]

The definition of a deontic update requires quite a bit of extra technical machinery to accommodate the conditional versions of the operators. We present them for information, though will only endeavour to briefly explain the various parts, instead we direct the interested reader to the original paper.[5] [33] Deontic relations in DUS are *reflexive* but not *transitive*, this is due to a technical problem related to the conditional nature of the operators.

6.6.1 Defining Conditional Operators

The link-cutting, of φ and $\neg\varphi$-worlds is done through the *reduction* function $\Downarrow$. The resulting state from a reduction is a state with all the links between states in which the action is true and states in which the action is false are removed from the requisite deontic relation.

[5] We use slightly different notation from the original in order to better suit the current presentation.

Definition 16. (Reduction)
Let σ be a promise state, and $i, j \in I$, $n \in \mathcal{N}$, such that $O^n_{ij}(\varphi|\psi)$ is a sentence of L^P_1.

$\sigma \Downarrow O^n_{ij}(\varphi|\psi) = \langle W, W_i, R^A_i, R^P_i, R^W_i \leq^n_{ij} -\{w_1 \leq^n_{ij} w_2 | w_1 \models \neg\varphi\wedge\psi \text{ and } w_2 \models \varphi \wedge \psi \text{ and } w_1, w_2 \in W\}, \preceq_i, V\rangle$

Before we can perform a deontic update, we must first take the transitive closure of the ordering. The transitive closure of an ordering can be taken by adding links $w_1 \leq w_3$ to the ordering for all $w_1, w_2, w_3 \in W$ such that $w_1 \leq w_2$ and $w_2 \leq w_3$. Taken as an iterative process, the fix point is the smallest superset of the ordering which is transitive as well as reflexive. [33]

Finally we define $pref_i$, as the function which whether or not the best ψ-worlds are also φ-worlds.

> The best ψ-worlds of W of σ satisfy φ if and only if for all worlds $w_1 \in W$ such that $w_1 \models \psi$ there is a world $w_2 \leq w_1$ such that, $w_2 \models \psi$ and for all worlds $w_3 \leq w_2$ we have $w_3 \models \psi \rightarrow \varphi$. [33]

We have $pref^n_{ij}(\sigma|\psi) \models \varphi$, as when the best ψ-worlds of σ satisfy φ in $\leq^n_{ij}$. For clarity in the following definition we suppress agent subscripts and normative force superscripts.

Definition 17. ($pref$)
Let σ be a promise state, and φ, ψ modal-free sentences of L^P_0.
$pref(\sigma|\psi) \models \varphi$ if and only if for all ψ-worlds $w_1 \in W$ there is a ψ-world $w_2 \leq_\psi w_1$ such that for all ψ-worlds $w_3 \leq_\psi w_2$ we have $w_3 \models \varphi$, with $w_1, w_2, w_3 \in W$. [33, 8-9]

With these functions in hand we now define the deontic updates $[O^n_{ij}(\varphi|\psi)]_i$. This is read as "agent i is obligated to see to it that φ if ψ is the case, for j under the normative force n".

Definition 18. *Deontic Updates*
Let σ be a promise state frame, ψ a modal-free sentences of L^P_0, $i, j \in I$, $n \in \mathcal{N}$, and φ a sentence of L^P_1 of the form $O^n_{ij}(\psi)$.

> If $pref(\sigma \Downarrow (O^n_{ij}(\alpha|\beta)|\beta) \models \alpha$, then $\sigma[\varphi]_i = \sigma \Downarrow O^n_{ij}(\varphi|\psi)$,
> otherwise $\sigma[\varphi]_i = \mathbf{1}$.

6.6.2 Working with Conditional Operators

With the conditional deontic operators, and the OSaaT principle, we have a solution for the problem of conditional promises. We model the obligation as **oblige**$(\alpha|\beta)$, which cuts the appropriate links in the agent's ideality relation. Then if we update, epistemically, with β the agent is left in a state where there are only β-worlds, and thus α is ideal in all of them.

6.6.3 Problems with the DUS solution

The DUS solution is both technically complicated and unmotivated. That is, while we do end up with a mathematically sound relation it is not clear that this is an appropriate end state philosophically.

When attempting to model the workings of real-world agents we must be clear on *why* a state is the desired end-state; the DUS solution is missing this. Further there is no, and likely cannot be, any motivation for the technical steps taken. Why would the ideality relation of an agent be non-transitive? Is it the case that we do wait until we encounter a specific scenario and then commit the mental gymnastics required to reform our ideal notions of the world to a relative transitive relation?

This technical solution does fix a problem that results from changing Veltman's update semantics into a single collapsed domain. In Veltman's semantics for default rules there is a separate domain for every φ. Continuing on this approach for deontic, and doxastic, attitudes is clearly too complicated. However the DUS solution seems to point that a single domain is perhaps too simplistic, as we must commit to several unmotivated technical moves in order to have a mathematically, but not motivationally, sound result.

For the current purposes, we leave the problem of modelling conditional obligations as a still open problem.[6]

[6] As noted by our predecessors, [37, 313].

Chapter 7

Areas of Expansion, and Future work

7.1 Deontic Epistemology

Alchourron and Bulygin discuss the difference between a logic of norms and a logic of normative propositions, i.e. whether there are logical relations between norms, allowing the development of a *deontic logic*, or whether there are no such relations and we must shift to a formal system modelling the propositions of said norms.

A normative proposition is *descriptive*, while the norms themselves are *prescriptive* in nature. This is illustrated by Alchourron's box metaphor.

> We may depict the difference between the descriptive meaning (normative propositions) and the prescriptive meaning (norm) of deontic sentences by means of thinking the obligatory sets as well as the permitted sets as different boxes ready to be filled. When the authority α uses a deontic sentence prescriptively to norm an action, his activity belongs to the same category as *putting something into a box*. When α, or someone else, uses the deontic sentence descriptively his activity belongs to the same category as *making a picture of α putting something into a box*. A proposition is like a picture of reality, so to assert a proposition is like making a picture of reality. On the other hand to issue (enact) a norm is like putting something in a box. It is a way of creating something, of building a part of reality (the normative qualification of an action) with the purpose that the addressees have the option to perform the authorized actions while performing the commanded actions. [1, 1]

While it is clear that promises are *prescriptive*, the act of promising generates new obligations (norms), the other side has yet to be addressed.

A promise is a form of speech act, though not limited to the verbal utterance 'I promise'. It results in the illocutionary and perlocutionary effects of changed deontic and doxastic states, of the promisor and promisee respectively. Speech acts also have epistemic effects, though generally a straightforward, and less interesting affair.

The utterance of a promise also informs both parties, and any others present, of the existence of the promise, and thus also of the entailed obligation. That is, there is an epistemic effect to promise-making. A promise is *both* normatively prescriptive and descriptive.

In deontic update semantics van der Torre and Tan model descriptive obligations by limiting the deontic operation to the actually known states W_i, while the prescriptive updates range across all worlds W, even if they are known to not be possible. This approach will not help in the case of promises, which are members of both sets, as the prescriptive update already subsumes the descriptive, $W_i \subseteq W$. The idea that a descriptive update only affects W_i is appealing, as W_i represents the epistemic state of the agent. However the update still only affects the ideality relation $\leq$ in DUS, which is the deontic state of the agent. Descriptive updates, the picture of the boat, are *informative* not *normative*. Thus the changes should be in W_i not $\leq$.

The class of puzzles examined thus far did not trade on the deontic epistemology of promising, but there are those that do, e.g. the problems of *breach*, *forgiveness*, and *absolvement*.

> *You are arriving at Schipol on Tuesday morning at 9:00am.* ***I promise to pick you up at the airport.*** *After you land you wait around at the airport until well after 10:00am, but I am nowhere to be found. You call my mobile and accuse me of* promise-breaking.

Let us model this situation in the standard manner. We start in the state σ. Then a successful promise is made.

$$\sigma[Prom^{f}_{ij}(\varphi)]_{ij}$$

I do not pick you up at the airport.

$$\sigma[Prom^{f}_{ij}(\varphi)]_{ij}[\neg\varphi]_{I}$$

While it is clear to the outside observer that the obligation has been broken, agent j does not have *knowledge* of the obligation, they only have a strong belief that i will do φ, $B^{S}_{j}(W_i(\varphi))$. However a strong belief in someone's actions does not necessitate that an obligation exists, e.g. I have a strong belief that the reader of this sentence will continue until the '.', but if this does not happen I have no normative claim against them.

For puzzles of this nature we must turn to a more developed system, one with *obligation tracking* within the agent's epistemic state. This gap is not limited to issues of breach, but also of any situation where there are post-promissory events (updates), e.g. normative clashes, cancellation, and normative evaluations. For

instance, the promisor cannot petition for absolvement without both parties having *explicit knowledge* of the obligation, not just the circumstance

In order to deal with these situations we need a more epistemically enriched system. A good candidate would be a DEL-like system, such as DEUL, [32]. In these systems the epistemic and preference operators exist in the basic language and the modalities can be nested. Such a system would still need to be enriched to appropriately model the obligation-creating-circumstances and doxastic operations, if we are to retain the current results, or model complicated puzzles involving both spheres.

7.2 Forgiveness and Punishment

Promissory breach is one example of the general idea of the 'latter half' of a promissory situation, i.e. situations which change/continue after the promise has been successfully made. Following breach, then, is the question of what *actions* are available to the promisee.

We can update with the breach of a promise, $[\neg\varphi]$, relying on the OSaaT principle. However the formal system is not strong enough to show what the actions available to either party are. The agent's possible actions are modelled by the modal alethic relation, and its unary operator C_i.

In order to model more complicated action/choice phenomena the basic language L_0^P should be shifted to a stit logic. At the same time, temporality would need to be explicitly introduced, likely in a branching time model. The OSaaT principle takes us quite far in showing promissory phenomena, but does not do justice to actually modelling specific promises. So we turn to action theory.

Such a move would allow for more complex puzzles that deal with the promisor's preconditions, (foregone) petitions for absolvement and the promisee's options post-breach.

We have dealt with concrete promises thus far, but not all promises are temporally fixed. The current system is ill-equipped to deal with other types, e.g. *open-ended promises*.

> *We are out at a café, and I have forgotten my wallet. You buy me a coffee.* ***I promise to buy you a coffee next week****. The following Monday we are out, and I do not buy you a coffee.*

Let σ be the original state. I successfully promise you that I will buy you a coffee in the future.

$$\sigma[Prom_{ij}^{f}(\varphi)]_{ij}$$

On Monday I do not buy you a coffee

$$\sigma[Prom_{ij}^{f}(\varphi)]_{ij}[\neg\varphi]_{ij}$$

We would not want to call this a broken promise, but we see the situation is modelled identically as the breach above. Open-ended promises require a more fine-grained notion of temporality than can be managed within PUS. For this

Normative Force	*Punishment*
Authority	Reprimand
Honour	Rebuke (Disappointment)
Promise	Restitution

Table 7.1: Punitive actions available to the promisee.

an action/temporal logic basis would be needed for our update semantics than our simple modal case.

Obligations are generated under different normative forces, and the available actions of the obligatee differ depending on the class of obligation. An example of normative-punitive mapping is given in 7.1.

7.3 Framework Clashes

There are other uses for PUS. As we can model the deontic and doxastic changes according to multiple frameworks, we can also do this *concurrently*.

We can model situations where one agent adheres to one of the promissory frameworks, and the second agent is a proponent of another. That is, with this system we can also model the clash between normative views on promises within puzzles, as discussed above.

$$[Prom^{e}_{ij}(\varphi)]_i[Prom^{c}_{ij}(\varphi)]_j$$

7.4 Conclusion

The primary aim of this essay was to explore the various philosophical frameworks that explain what a promise is, and how obligations are tied to the speech act, including a new framework that proposed to solve the problems left by the two main families of promissory account.

We started, in chapter two, with a discussion of promises as obligation-creating speech acts; noting that circumstances must be just so in order for a promise to successfully take place. Clear differences involved the appropriate promisee, whether or not the act allows for agent-reflexive obligations, and the normative force that grounds the resulting obligations. The chapter was rounded out with a brief survey of promissory frameworks, broken into the families of *conventional* and *expectational* accounts, ultimately married together in our *conventional expectations* hybrid approach to promising.

Chapter three continued the exploration into promising by first discussing the cardinality of the audience of a promise, and then comparing promises to other obligation-creating speech acts, noting their similarities and differences. We introduced a set of normative force indicators, $\mathcal{N}$, to indicate these differences.

We developed a logical framework to model these frameworks, with an eye to formally model the differences that arise under various puzzling circumstances.

Promissory Update Semantics (PUS) is a system that combines DEL-style doxastic updates with the deontic update semantics developed by van der Torre and Tan [33]. An advantage of the update semantics approach is that it is clear what the agents states are. That is, W_i is the epistemic state of i, the relation $\preceq_i$ their doxastic state and their deontic relations are represented by the binary relations $\{\leq_{ij}^{n}\}_{n\in\mathcal{N},j\in I}$. Noting that one can have obligations both to φ and $\neg\varphi$ if they occur across normative forces, or agent pairs. Chapter four finished with the formal modelling, in PUS,of some of the results from Chapter three as well as the definition of 'a promise' under the three promissory accounts.

Chapter five leveraged PUS, and its modelling of promises, to deal with puzzles in promising. These puzzles show the applicative differences between the promissory frameworks, adjusting intuition to promissory answer.

Up until this point we treated promises as simple atomic propositions, e.g. "pick you up at the airport" as φ. It is clear that we can promise more complex utterances. Chapter six tackled the problem of the logical form of promises. The operative question: What connectives can be combined with promising and retain semantic coherency? We saw with negations that a negation in front of a promise, a *negated promise*, has no mathematical counterpart which matches our intuition that one cannot commit such a speech act. Conjunctive promises have the minor problem of framing. While it is not clear what a disjunctive promises would be, unless a veiled threat. The sticking point, that we are still left with, is the *conditional promise*. And this is because promised acts are *temporal*. For all other logical forms of promising the temporality is of one act in the future, so our temporal gloss of $G_i(\varphi)$ does not impede us. With conditional promises the antecedent represents some future state, or action, which must first be reached before the consequent promised act is obligated. This means that we need a more refined temporal approach in order to model conditionals. Such an approach exists in DUS, but what we gain in the mathematical, we lose in the philosophical.

In Chapter seven we considered other classes of promissory puzzles that the framework could be extended to deal with. For example, temporality in the form of problems of forgiveness and punishment we must layer our update semantics on an action-temporal logic. Epistemically, promises are unique in that they don't just represent *prescriptive norms* but they are also *descriptive norms*. That is, a promise not only creates an obligation but it also informs of said obligation. Modelling this would require an extension into a more DEL-like epistemic system.

While the system presented here does not answer all the problems encountered when dealing with promissory phenomena it does represent a promising start.

Bibliography

[1] C. Alchourrón. Philosophical foundations of deontic logic and the logic of defeasible conditionals. In J. J. Meyer and R. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*. John Wiley & Sons, 1993.

[2] J. L. Austin. *How to do Things with Words*. Harvard University Press, 1962.

[3] A. Baltag. Topics in DEL: Course notes. Slides, University of Amsterdam, 2012.

[4] S. Bankins. Expectations, obligations or promises? A conceptual review and revision of the beliefs comprising the psychological contract. *24th Annual Australian and New Zealand Academy of Management Conference*, 2010.

[5] S. Beller. Conditional promises and threats - cognition and emotion. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, pages 113–118, 2002.

[6] S. Beller and H. Spada. The logic of content effects in propositional reasoning: The case of condtional reasoning with a point of view. *Thinking & Reasoning*, pages 355–378, 2003.

[7] S. Beller A. Bender and G. Kuhnmnch. Understanding conditional promises and threats. *Thinking & Reasoning*, 11:209–238, 2005.

[8] Guillaume Aucher Guido Boella and Leendert van der Torre. Prescriptive and descriptive obligations in dynamic epistemic deontic logic. In *AI Approaches to the Complexity of Legal Systems,*. 2009.

[9] E. Brake. Is divorce promise-breaking? *Ethical Theory and Moral Practice*, 14:23–29, 2011.

[10] W. R. Carter. Grice on promising on condition. *Analysis*, 30:31–32, 1969.

[11] R. S. Downie. Three accounts of promising. *The Philosophical Quarterly*, 35:140:259–271, 1985.

[12] R. G. Durrant. Promising. *Australasian Journal of Philosophy*, 41:1:44–56, 1963.

[13] G. R. Grice. *The Grounds of Moral Judgement*. Oxford, 1967.

[14] A. N. Habib. *The Authority Theory of Promises*. PhD thesis, University of Arizona, 2006.

[15] A. N. Habib. Promises to the self. *Canadian Journal of Philosophy*, 39:537–557, 2009.

[16] Allen Habib. Promises. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2008 edition, 2008.

[17] D. Hume. *A Treatise of Human Nature*. Oxford: Clarendon, 1896.

[18] N. Kolodny and R. J. Wallace. Promises and practices revisited. *Philosophy & Publics Affairs*, 31:119–154, 2003.

[19] A. H. Kurji. The speech act of conditional promising. Course paper, University of Calgary, 2008.

[20] M. Migotti. All kinds of promises. *Ethics*, 114:60–87, 2003.

[21] D. Owens. A simple theory of promising. *Philosophical Review*, 115:51–76, 2006.

[22] Peter Pagin. Assertion. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2008 edition, 2008.

[23] M. H. Robins. The primacy of promising. *Mind Association*, 85:321–340, 1976.

[24] D. M. Rousseau. Schema, promise and mutuality: The building blocks of the psychological contract. *Journal of Occupational and Organizational Psychology*, 74:511–541, 2001.

[25] T. M. Scanlon. *What We Owe to Each Other*. Belknap Press, 1998.

[26] J. R. Searle. How to derive "ought" from "is". *The Philosophical Review*, 73:43–58, 1964.

[27] J. R. Searle. *Speech Acts: An Essay in The Philosophy of Language*. Cambridge University Press, 1969.

[28] Kieran Setiya. Intention. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2011 edition, 2011.

[29] H. Sheinman. Promise as practice reason. *Acta Analytica*, 23:287–318, 2008.

[30] A. Smith. *The Theory of Moral Sentiments*. Oxford, 1976.

[31] J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 14, 2004.

[32] J. van Benthem and F Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 14, 2004.

[33] L. van der Torre and Y. Tan. An update semantics for deontic reasoning. In H. Prakken P. McNamara, editor, *Norms, Logics and Information Systems*. "IOS Press", 1999.

[34] R. van Rooij and M. Franke. Promises and threats with conditionals and disjunctions. In Grewendorf and E. Zimmerman, editors, *Ten Years After*. forthcoming.

[35] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.

[36] T. Yamada. Logical dynamics of commands and obligations. In H. Takeda T. Washio K. Satoh and A. Inokuchi, editors, *New Frontiers in Artificial Intelligence, JSAI 2006 Conference and Workshops*, pages 143–146. Springer, 2007.

[37] T. Yamada. Logical dynamics of some speech acts that affect obligations and preferences. *Synthese*, 165:259–279, 2008.

[38] Tomoyuki Yamada. Acts of promising in dynamified deontic logic. *New Frontiers in Artificial Intelligence, JSAI 2007 Conference and Workshops*, 2008.