A static theory of promises

Bergstra, J.; Burgess, M.

Publication date
2008
Document Version
Submitted manuscript

A static theory of promises

Jan Bergstra*
Section Software Engineering, Instititute of Informatics
University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam
The Netherlands

Mark Burgess†
Faculty of Engineering
Oslo University College
PO Box 4 St Olavs Plass 0130, Oslo
Norway

October 18, 2008

Abstract

We discuss for the concept of promises within a framework that can be applied to either humans or technology. We compare promises to the more established notion of obligations and find promises to be both simpler and more effective at reducing uncertainty in behavioural outcomes.

1 Introduction

This paper is about the usefulness and importance of promises as a concept. For something that is so abundant in daily life, the concept of a promise has been given only a light superficial treatment in academic literature, and then only in the areas of philosophy and economics. Promises give way more often to the idea of obligations, indeed many authors quickly do away with promises and replace them with apparent obligations as if this were a necessity. No doubt this tendency originates in historical matters, but it is unfortunate as promises have a plausible and practical value both in their mundane meaning for humans and in an extended interpretation that can apply to machines and other inanimate objects.

The concept of a promise is not a difficult one – in fact it is a much simpler concept than an obligation – but it requires some care. In this paper we present our understanding of promises an explain why there are both practical and theoretical advantages to their use over obligations. To make our case unambiguously, we ask the forbearance of readers in presenting the fundamentals and motivations at some length.

2 A motivation

For surely all readers the notion of a promise will be quite familiar. Our experience of promises might be both positive and negative, for many promises that are made are never kept. Below are examples of the kinds of statements we intend to call promises. Let us

*Jan Bergstra acknowledges support from NWO reasearch grant Symbiosis.
†Mark Burgess is supported by the EC IST-EMANICS Network of Excellence (#26854)

begin with everyday statements and progress gradually to the kinds of abstract promises that we would like to use in a variety of technical scenarios.

- I promise you that I will walk the dog.
- I promise you that I fed your cat while you were away.
- We promise to accept cash payments.
- We promise to accept validated credit cards.
- I promise to lock the door when I leave.
- I promise not to lock the door when I leave.
- We promise to wash our hands before touching the food.

These examples are quite uncontroversial. They are easily found in every day life, spoken by humans or posted on signs. We now want to argue that it is useful to extend the notion of promises to inanimate objects that have been designed or programmed to behave in a certain manner. This is not a very large step, but it is easier to make with some examples.

Consider the following promises that might be made in the world of Information Technology:

- The Internet Service Provider promises to deliver broadband Internet at a fixed for a fixed monthly payment.
- The security officer promises that the system will conform to security requirements.
- The support personnel promise to be available by pager 24 hours a day.
- Support staff promise to reply to queries within 24 hours.

Again these are straightforward promises, which could be described further to be more specific. The final promise could also be restated in more abstract terms, transferring the promise to an abstract entity: “the help desk”:

- The company help-desk promises to reply to service requests within 24 hours.

This latter example illustrates the way that we transfer the intentions of promises to ‘entities’ that we consider to be responsible by association. It is a small step from this transference to a more general assignment of promises to individual components in a piece of technology. For example, we can document the properties of the following tools and technologies in the spirit of this argument:

- I am a meat knife and promise to cut more efficiently through meat.
- I am a logic gate and promise to transform a TRUE signal into a FALSE signal and vice versa.
- I am a variable that promises to represent the value 17 of type integer.
- I am a command line interpreter and promise to accept input and execute commands from the user.
- I am a router and promise to accept packets from a list of authorized IP addresses.
- I am a compliance monitor and promise to verify and automatically repair the state of the system based on this description of system configuration and policy.

- I am a high availability server and I promise you service delivery with 99.9999% availability.

From these example we see that the essence of promises is quite general. Indeed such promises are all around us in everyday life, both in mundane clothing as well as in technical disciplines. Statements about engineering specifications can also profitably be considered as promises, even though we might not ordinarily think of them in this way.

When an electronics engineer looks in a component catalogue and sees 'resistors' for sale promising to have resistance of 500 Ohms to within a tolerance of 5%, we do not argue about who made this promise or whether the resistor is capable of independent thought. The coloured bands on the component are a sufficient expression of this promise, and we accept it by association. By this reasoning, we propose that the concept of a promise should be formulated in way which allows for all of these uses.

The value of this association is that promises are things that we use to form *expectations* of the behaviour of all manner of things. Such expectations contribute to reducing our *uncertainty* about their behaviour, and this can apply as much to technology as to humans. We therefore take it as given that the concept of a promise is a useful one and consider next how one can formalize promises in the simplest and least assuming way.

3 The concept of a promise

Because of their overriding ubiquity, and practical importance, one would like to have an account of promises that captures their key properties and explains related concepts such as commitment, obligation and intention. There is a surprising lack of discussion about the meaning of promises in the literature as far as we can tell. Although the concept or its relatives have been mentioned in such diverse areas from logic, law and philosophy to economics, information science and computing, there is no agreement on what constitutes the semantic content of the terms or if there is even more than a tacit relationship between promise, commitment, obligation etc. The most attention has been given to the concept of *obligations* especially in the area of deontic logic. We believe on the other hand that the philosophical implications of promises are far wider than is generally assumed and that there is both a need and a practical importance to clarify them once and for all. Indeed, we will show that the concept of a promise is far simpler than that of an obligation.

Atiyah [1] suggests that any promise leads to an obligation to keep that promise that is motivated by the threat of tit for tat reprisals. Reciprocation is thus coupled to the idea of promises immediately, which seems to hop over fundamental definitions directly to a discussion of the economics of keeping promises. The obligations are to avoid injury and to reciprocate goodwill. It might be discussed whether incentives are the same as obligations. Atiyah points out however that promising something cannot be necessarily used to create obligation at will. Promises might cause obligations but they can also represent obligations that already exist, i.e. to show commitment to an existing obligation to pay the price of something. e.g. I promise to pay the bearer the sum of 1 pound (in gold). This is only an existing admission of moral obligation. Atiyah maintains, plausibly, that the motivation for promising has changed throughout history. When people make promises, their intentions are culturally bound. Thus a Victorian gentleman's conception of a promise might not fit with that of a present-day child who promises to be home in time for dinner.

Cartwright takes Atiyah's view and asks what might be the point of promises if not to generate the assumed obligation[2]. Why do people bother to make promises about things to which they are already obliged? His answer includes the idea that it is a face-saving measure: to mitigate their humility, suggesting that an obligation is interpreted as a kind of attack or levy of force? Alternatively, perhaps the obligation to keep one's promises weighs heavier than the original obligation (I promise you my word as a gentleman not to kill you, even though the law says I am forbidden). Referring to Fried[3], Cartwright points out that

the economics of contractual tit-for-tat suggested by Atiyah is tied to promises and not to the obligations they might confer.

The idea that promises are an economic driver of contracts or agreements as bilateral exchanges of promises is continued in the work of Gilbert [4]. Then Carrillo and Dewatripont have argued that promises can best be understood as a market mechanism for reducing the uncertainty in a moral-hazard game [5]. This work does not seem to have been pursued. Does a promise increase the likelihood of voluntary cooperation? A number of other works mention the concept of promises in the context of game theory also. In these, the concept of a promise is tacitly assumed to be related to the probability of choosing a particular game strategy.

More recently, a different motivation for promises was introduced by Burgess in the context of distributed management [6]. Rather than focusing on morals or even economics as the principal motivator, Burgess uses the promise as a measure of 'voluntary cooperation' as a way of resolving fundamental problems with logics of obligation for determining system behaviour. Voluntary cooperation is seen as a way of simplifying constraints and avoiding many-worlds paradoxes. He pursues the argument further by emphasizing the role of autonomy of the parts, and argues for a 'promise theory' in which every component in a system that can have unique information or independent action should be viewed as axiomatically autonomous [7]. Any cooperation or even subordination of the parts that comes about in an organized system must then be understood as the result of 'honouring' purely voluntary promises to do so. Burgess argues that no matter what one believes about the power to oblige (even soldiers can refuse to follow orders), voluntary cooperation can be used as a pragmatic engineering methodology for mapping out the complexity of a control problem in a way that is invariant with respect to centralization or decentralization of systems.

In computer science, particularly the field of Multi-Agent Systems the concept of *commitments* has been used for some time [8, 9]. It has been suggested that promises and commitments are the same. However, we shall show that this is not the case. More seriously, the sense in which the term commitment is used in such discussions is more stylized than purposely considered and can only benefit from the discussion in this paper.

4 A model of the structure of a promise

Consider the following intuitive idea of what a promise might be: *A promise is an announcement of fact or behaviour by a promiser to a promisee, observed by a number of witnesses (referred to as the scope of the promise), whose outcome has yet to be assessed.*

The promiser and promisee are both assumed to be 'agents', i.e. humans or inanimate objects to which we attach identity in the story of promises. This general description fits the examples that we have already given and gives some clues as to the constitution of a promise, but it also opens up a number of questions that need answering. Already we can see that this apparently basic definition rests on a number of assumptions: that we can observe the outcomes of behaviours and that the outcome of a promise is clear at some single moment of time in the future, to be measured and verified by an observer. A full account of this might include a theory of measurement, but we wish to avoid this level of detail as it binds us to too many details that have nothing directly to do with the issue. Let us instead try to understand to essential characteristics for promises and consider what distinguishes a promise from related matters, such as obligations, commitments and other terms.

The model world in which we formulate promises must have the following characteristics.

- There must be agents in order for promises to exist.
- There must be a promiser (or source agent).

- There must be a promisee (or recipient agent) which might be the same as the source.
- There must be a body which describes the nature of the promise.

We might summarize these attributes with a notation as in ref. [6]:

$$promiser \xrightarrow{body} promisee \tag{1}$$

- We can leave the body unspecified, but it must consist of a quality (a type, topic or subject for the promise) and a quantifier (which indicates how much of the realm of possibility for that subject is being promised). For example: promise quality: "travel to work", promise quantity "on Monday and Friday each week".

Finally, what is implicit in the above is that a promise requires the transmission of a message, or at least documentation in some kind of physical form, e.g. a speech act, or a written statement, else it cannot be made known to anyone except the promiser. A promise must therefore have documentation that is made intentionally or otherwise.

What then is a promise before we write it down? We shall refer to this as an *intention*. An intention is the basic formulation of a course of behaviour, which is made internally by (or on behalf of) an agent. When an intention is made public, it becomes a promise. If an intention is documented or leaked in some way then anyone has a right to assume it is a promise.

We take it as given that there has to be a source for every promise. A promiser does not have to reveal its identity of course, so witnesses to the promise might not know its source e.g. consider the anonymous threat. There is no reason to deny the existence of a source however. The lack of such information about a promiser is simply a defect in the knowledge of the receiving agent, but one would normally prefer to assume a consistent picture of promises and infer the existence of an anonymous promiser. This justifies our postulating the source.

5 Promises are documented intentions

A key characteristic of a promise is that it documents an intention, so let us explore the idea of intentions in more detail. Intentions turn out to be a lowest common denominator for all of the concepts discussed in this paper and thus have a special importance.

Since promises involve communication we require a notion of the spread of information amongst the agents. We use the term *scope* (as used in computer science) for this. A scope is simply defined to be a set of agents. For example, the scope of a promise would typically refer to the promiser and a list of witnesses to the promise, e.g. those who heard to utterance or those who saw the written document.

5.1 Intentions

The realm of all possible formulations about behaviour is covered by the concept of intentions.

Term 1 (Intention) *A description of possible behaviour that can be contemplated by an agent and be brought to realization.*

The components of an intention are as follows: a source agent who formulates the intention, a target agent if the intention is directed at a potential subject, and a body which explains the quality and quantity of the intention (see fig. 1). Only the source of an intention knows about the intention, i.e. the scope of an intention is the source only. There are no witnesses.

Now we must be careful: the set of all possible intentions should be distinguished from actual instances of intentions selected by an particular agent. We shall sometimes use the phrase "possible intentions" to mean this full set of abstract entities to emphasize when we wish to signify a general description of behaviour rather than an individual agent's decision.

An intention is not announced by the agent holding it to any other agents. Indeed, we may now *define* any intention that is announced to be a promise. Conversely we notice that any promise that has not been announced is merely an intention. Some intentions are desirable while others are absolutely undesirable and an agent might never choose them, yet they are possible intentions nevertheless. The fact that such behaviours can be intended is enough for them to qualify as intentions ("possible intentions").

Intentions must always be thought of as belonging to a specific agent. Those intentions which are actual plans of the agent are called its commitments. Other phrases for a commitment that we may use are: intended intention, or real intention.

Due to the static nature of our account we pay no attention to the process by which an intention might become a commitment or vice versa.

5.2 Promises

A promise is the physical publication of an intention within a certain scope. This suggests that there must be some agent to observe the promise and its outcome which in turn requires the expensive notion of a theory of observation so we shall tackle this issue separately (see section 7).

Term 2 (Promise) *A promise is an intention that has been documented within a scope that goes beyond the promiser.*

Promises thus have scope. Formally intentions also have a scope, but the scope of an intention held by an agent is by definition limited to the agent (source) itself. An intention could be leaked deliberately (e.g. to the press, in order to influence someone). This might be a form of leverage, or an attempt to impose an obligation on some party in scope. However, at the instant an intention expands in scope to encompass more agents it becomes a promise. A so-called letter of intent, for example, is a promise rather than merely an intention.

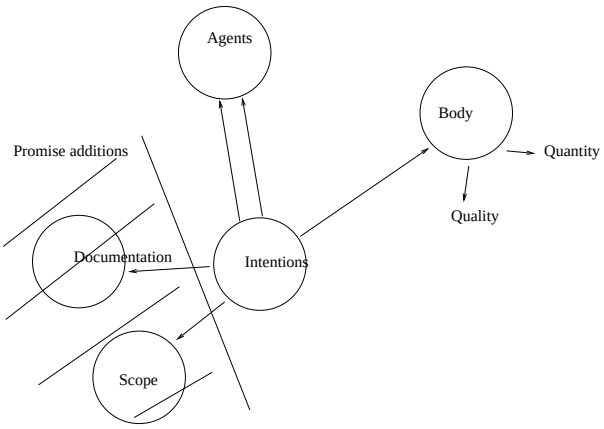

Figure 1: Intentions and their structure.

The time aspect of promises presents further challenges. Intentions can become outdated by events. An event which is found to fulfill and intention documents the implicit promise, since one must admit to the intention in a wider scope. Conversely, the documentation for a promise does not have to last for ever; if documentation of a promise disappears

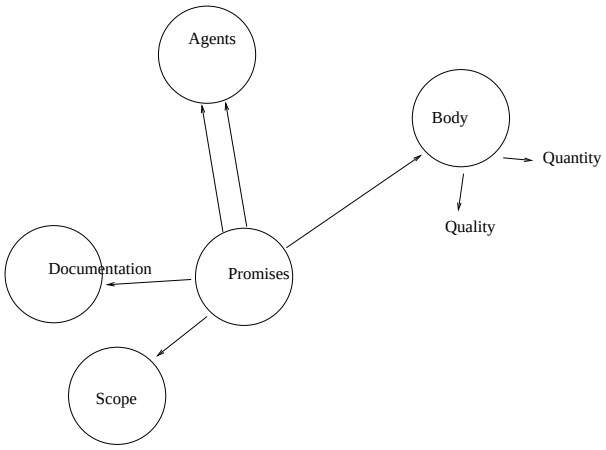

Figure 2: Promises are intentions with documentation and an audience.

completely, it reverts to being an intention. A promise to oneself is merely an intention, unless it is documented.

The distinction between the promisee and the scope of the promise is key to understanding promises. Suppose someone intends to arrange a surprise birthday party for their friend. Initially this is an intention. The intention is written in a diary or mentioned to a third party and it becomes a promise. The target is not in the scope of the promise however, so the promise remains unknown to the jubilant. However, suppose that before telling anyone else, the promiser destroys all evidence of the promise by tearing out the page of the diary, effectively withdrawing knowledge of the promise, then the promise reverts to being just an intention. But as long as knowledge of the promise remains "out there" in the world, it remains a promise that has been made.

5.3 Obligations

Having explained intentions and promises, let us now try to describe the notion of obligations in the same manner. The intuitive notion of an obligation seems straightforward, but it proves to have difficult properties. We might try to think of obligations in a straightforward way, for instance: *an obligation is an intention that is perceived to be necessary by an agent*. This certainly captures some of the characteristics that we understand by the term, but it also leaves many questions unanswered: is the feeling of the necessity voluntary or forced, a matter of survival or simply an authoritarian convention?

Unlike a promise, an obligation might be self-imposed or externally imposed. An obligation falls into the category of (possible) intentions, so it must have source, a target and a body, and the body must have a quality and a quantity. The source and target are now somewhat difficult to understand however.

Beyond this, we shall not attempt to define obligations more carefully in this paper. We shall merely state some assumptions about them.

An obligation can be imposed by external conditions, e.g. by the expected behaviour of external agents, by laws threatening sanctions etc, or it can be self-imposed by codes of personal behaviour which an agent holds to be *necessary*. But this imposition suggests the action of a force which attempts to induce a commitment in another agent (or itself). An obligation is an intention ("possible intention") which may or may not have the status of a commitment. In any case the agent is aware of any compelling reasons to include the intention in the portfolio of commitments, either from within itself or without due to external forces.

It seems natural then to refer to the source and target of the induced intention as being

the agent in which one attempts to induce the intention, and the recipient of the intention respectively. However, the source of the obligation itself might not be an agent at all, but merely a set of external conditions, norms, experiences or other information acquired by the agent that lead to a perceived priority.

Note again that even 'forced' behaviour can be classified under the realms of (possible) intentions since all behaviour can be intended. Again, we emphasize that this does not imply that a coerced agent holds the intention that is being forced upon it. Nor does it say anything about whether the agent is able to resist the force or not, or whether it matters if an obligation is self-imposed or externally imposed.

The notion of an obligation immediately seems far more complicated than an intention or a promise and does not seem to be close to the notion of either promises or intentions. We hold that obligations are far from being a reliable tool for ensuring compliance. If a law-giver wanted to ensure the compliance of an agent, a better strategy would be to obtain a promise from the agent, and to convince it to view the intention as a *commitment* since the law-giver could never know whether the agent had indeed committed to the body of the obligation.

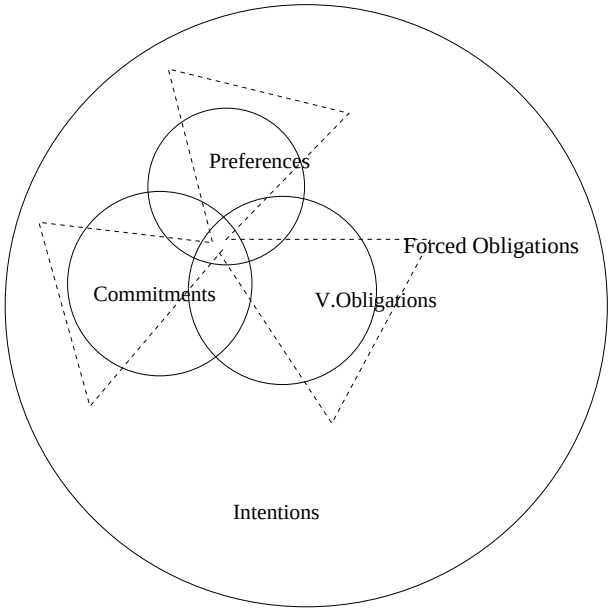

Figure 3: The voluntary cooperation universe and the ranking of intention. Externally applied obligations can overlap with any of the categories of formulatable intentions.

To study the idea of compliance further let us return now once again to promises. It is clear that promises and obligations cannot be simply related (as some promises might be deceptions) so let us explore deceptions in more detail.

5.4 Deceptions – non-intended promises

Understanding deceptions (or lies) is also an important step in clarifying the relationship between intentions and promises, because it is possible for an agent to have two different intentions in play at the same time: a commitment and an announced intention (i.e. a promise) which are not compatible. Incompatibility means that striving for both intentions simultaneously is fruitless because their realizations cannot be combined.

In a deception, there is always a source and always a target and the target cannot be the same as the source, as an agent cannot (intentionally) deceive itself. Furthermore, we maintain that the target of a deception must be in scope, so there must be a physical

documentation and hence a deception necessarily involves a promise and not merely an intention.

Term 3 (Deception) *A deception consists of two intentions: a documented intention (i.e. a promise) and a non-documented commitment, which are incompatible.*

The non-documented commitment will also be called the hidden intention.

In a deception the hidden intention is more important than the witnessed one one and we might refer to it as the dominant intention. This simply expresses that it is a commitment while the promise contains merely a "possible intention". It is the *real* intention of the agent ("intended intention"), while the intention in the promise can merely be described as *non-real*. If the dominant intention should be rescinded, a deception will revert to being a promise, but this is only known to the source.

5.5 "Non-deceptions"

A deception is the augmentation of a promise with an *incompatible* intention. We should like a name for the augmentation of a promise with a positive intention. We might call this a promised commitment, or intended promise. From these slightly strained terms, we can now appreciate why the concept of a promise is in fact so important. A promise is simply a promise (the documentation of an intention), regardless of what lies behind it. Any internal priorities or considerations are hidden from the view of other agents and cannot be observed. Thus, promises are an independently important concept because we can (indeed must) talk about promises without discussing the basis on which they are made.

When a promise is made, we are neither required nor able to confront the truth or falsity of the promise. Indeed, as soon as we ask such questions, new issues such as trust and a plethora of other subjective issues come into play. Such issues are probably un-resolvable in a logical sense. However, what we assume is here that no matter how trustworthy a promise might be, it can increase or decrease our certainty of a promised outcome and thus it bears an *influence*. The matter of assessing the promise can be very complicated and uncertain and we shall not attempt to discuss this here in any depth.

6 Ranking of intentions

The foregoing discussion of deceptions suggests the existence of a ranking function which induces a partial ordering onto the intentions that are referred to or in play at any given moment. There are intentions one intends to invoke (i.e. that one commits to), intentions one prefers, intentions one feels obliged to intend, and finally deceptions which one intends to not honour the intention documented for a wider audience.

6.1 Commitment

To commit to something is to make a decision in favour of it. The issue of commitments is therefore about the favourization of intentions. Commitment is a personal decision and has nothing to do with physical representation or communication, thus the issue of commitment precedes any discussion of promises. A commitment has a source, a target and a body, i.e. it is an intention. Like an intention it has no physical representation and does not have a non-trivial scope.

Once an intention becomes a commitment we often assume that some point of no return has been passed in the act of committing (deciding) about the particular intention. i.e. adding the intention to a list of commitments. For example, in a game of chicken in which two cars drive towards each other to see which one will swerve off first, a driver has committed to not swerving when the decision to not back down has been made[10]. This might have certain irreversible consequences, but it is difficult to generalize the idea of

irreversibility in examples of this kind. What commitment essentially boils down to is the elevation of some intention beyond an arbitrary threshold. In other words, in the universe of intentions there is a subset of these which we may call commitments.

Term 4 (Commitment) *Commitments are intentions that we are committed to. We may call them intended intentions, or equivalently real intentions, intentions that we hold, or committed intentions.*

6.2 A partial ordering

A given agent at a certain moment of time ranks intentions by applying to them a *partial ordering*. An intention is considered higher in this ordering if it is closer to a commitment. Commitments are at the top of this ranking; at the bottom are those intentions which are incompatible with commitments. The latter are termed negative intentions or contrary intentions. In between these extrema lie voluntary obligations, and involuntary obligations, and these will in general intersect with all other categories(see fig. 3).

7 Assessments

The notion of whether promises are kept or not is central to their sustained usefulness in society, thus we need to make mention of how this comes about in a theory of promises. It would be easy to go overboard and delve into the complexities of observation and measurement to provide a satisfactory answer but that is not in the spirit of this paper. We seek instead a simpler notion which is at the same level of abstraction as the concepts of promise and intention that we have introduced thus far. We call this the concept of *assessment*.

Term 5 (Assessment) *An assessment is a subjective statement made by an agent about whether the intentions of itself or of another agent were fulfilled.*

Our notion of an assessment is more generic and less quantitative than a verification. It is both subjective and not *a priori* linked to observation. In a static theory of promises and intentions the existence of intentions as well as the value of assessments is linked to state parameters like time. Thus, for an intention of agent A, in existence at time t, it may be the case that agent B's assessment, made at time $t' > t$ is positive (or negative).

At this level of description, we need not say any more about it than this. The reality of whether promises have been kept through specific actions is neither here nor there in the world of politics and to some extent economics. What is important is how a witness to the promise assesses the outcome of the promise. Such an assessor may or may not feel obliged to assess an outcome in a particular way, might promise to conform to certain criteria, and so on. What matters is only the assessment, which might or might not be rationally obtained. We believe that this is a fair model of the world in which we live.

An assessment is a supposed outcome relative to some method of assessment. Assessment involves a variety of possible routes to inference, i.e. there are different kinds of assessment. This is a subjective issue, but this should not be viewed as a weakness of our theory: it is a true feature of the subjective nature of individual assessment.

Some example assessments are shown in table 1. We see that assessments are quite sensitive to physical representation of the promise. Once again the notion of representation (or documentation) is a key to the importance of a promise as a concept.

8 The value of promises – relativity

Promises are valuable to agents, because they help reduce uncertainty and because their outcomes could be beneficial if they become certain. Because certainty is key, a promise

Promise	Representation	Scope	Assessment
Fed the cat	Speech act	Those who heard the promise	Either did or did not.
Credit card accepted	Action	Visitors to the store	Either did or did not.
Response in 24 hours	Contract	Signing parties	Reply received in time or not.
$var = value$	Source code	Readers of the code / compiler	Syntax ok. Value in range. etc
$var = value$	Object code	Execution engine	Ok to execute, exception, etc.

Table 1: Promise assessments

is worth nothing unless there is trust. Zero trust makes promises worthless. Trust might be based on a history of keeping promises or, in our terminology, on a history of positive assessments about a succession of promises. So there is a symmetry between trust and promises that must be broken to solve the dilemma.

If there is trust, a promise about future behaviour does not need to be perceived as an obligation on the promiser but as an indication that best effort will be respected. If a given agent X does not trust the promiser however, it might assume that the promise implies an obligation on the promiser. This perception of obligation by X does not of course imply an obligation perceived by the promiser. There is a fundamental subjectivity in these perceptions.

The value of a promise is an expectation of the eventual benefit. Suppose, then that A promises B 400 dollars per year. B promises to wash A's windows at this price. Both are satisfied with the value they get from this arrangement and prefer not to question it too much as this could unleash all kinds of consequences. Observer C can see that the values are quite mismatched, or that A is getting a poor deal by its judgement, but C also cannot deny that the relationship is stable because both A and B are happy.

The value of promises may be questioned by those who consider promises as a concept secondary to obligations. If one views obligations as the primary concept, the value or importance of promises unavoidably shrinks. We shall now survey advantages and disadvantages of obligations as an alternative cornerstone of a theory of promises. The discussion will be somewhat asymmetric because we will not base our comparison on a proper definition of the concept of an obligation (which we cannot fathom). Suffice it to say that for some agent A an obligation is an intention ("possible intention"), which has been elevated to the status of an obligation, whatever the consequences of this status may be.

9 For and against the primacy of obligations

Obligations are discussed extensively in the literature whereas promises have been ignored. By the sheer weight of tradition, obligations dominate discussions of behaviour.

9.1 In favour of obligations

1. Some people might think that a promise is an obligation because it seems to create one, and might therefore be considered equivalent to that obligation.

2. Obligations are a well known concept from deontic logic. There is an advantage to to reducing the less well-known concept of promises to one that has been studied for more than fifty years.

3. Obligations have a formal status in state laws and regulations. There is no such public body of promises.

4. Many obligations give rise to promises which occur in the process of fulfilling an obligation. e.g. the cat must get fed while owner is on holiday, the owner is obliged to get the cat fed (by law forbidding cruelty to animals). A friend promises to help in the fulfillment of the obligation.

9.2 Against obligations

1. If a future promise (e.g. the promise to feed the cat in the future) is in fact a deception then this is falsified the necessity of a relationship between promises and obligations. In other words, all promises cannot be obligations because some promises can be deceptions and these cannot be understood as such.

2. All descriptions of deontic logic are fraught with logical difficulties.

3. Not all promises are about future actions, so there cannot be an implied obligation for all promises. e.g. I promise that the cat got fed. Indeed the owner might actually be displeased that the cat was fed if it was supposed to be dieting.

4. In law, it is true that there is a dissimilarity between promises and obligations. They are quite different entities. Obligations may cause promises and promises may cause obligations, but promises have a physical representation in space and time, whereas obligations do not. Obligations are at a different level of abstraction altogether. Moreover, the international monetary system can be viewed as an example of a de-facto standard promise – the promise to redeem the value of money on a voluntary basis. It is also well known by law-makers that laws are only observed about issues that most people will basically keep to voluntarily, thus the power of obligation may be a fictitious one, based on de-facto promises.

5. Promises are made on a voluntary basis. For obligation however, the concept of voluntarity is almost irrational. In any case it might be voluntary to imply an obligation on someone else, but engaging in a promise you may face an involuntary obligation or a voluntary one. Voluntary is therefore natural for promises but is quite problematic for obligations.

6. Promises announce positive extensions of user behaviour, obligations are a negative constraint on the degrees of freedom of the obliged party.

If one would choose between promises and obligations, it seems abundantly clear that promises are the simpler concept. Moreover, the concept of a promise seems more natural in the technological world: since computers cannot feel ethical responsibility, the reduction to promises to obligations seems to be neither philosophically satisfactory nor technically correct. Moreover, there are some behaviours one cannot oblige (empty the ocean with a sieve). These can be promised, even if the promises are clearly deceptions.

So far we have argued that promises are different, simpler and can be analyzed independently of obligations. There is one more point that is of principal practical importance. Promises are *local* constructions, whereas obligations are *non-local*. The source of a promise is localized in a single entity that has all of the information and self-control to be available to resolve conflicts and problems with multiple promises. The sources of obligations however are distributed amongst many individuals and the obliged party does not have the access to resolve the conflicts without maintaining a voluntary dialogue with all of these multiple parties.

From a practical viewpoint, obligations are simply less effective at reducing uncertainty because they tend to increase uncertainty not reduce it. Indeed, obligations can be inconsistent, but promises cannot. More precisely: consistency of promises is a matter that can be verfied at the level of sources only. Promises made by different agents cannot be inconsistent.

Preferably then one would not use obligation as a coordination principle if a mechanism based on promises can be used instead. Promises are simply more trustworthy. A collaboration based on promises works better if one has trust. In a world of obligations however, trust is meaningless because one has only a presumed outcome.

10 Future work

We have restricted our attention to static scenarios in this work, as the matter of change is a complex one. An account of how promises arrive, persist and are removed again is forthcoming. Some work has already been done in this area, however[11, 12] but scope for embellishment is vast, as is the number applications for the concept of promises. In the latter reference, the matter of *organization* is related to promises, as a form of cooperation between individuals or autonomous agents.

Although we have avoided describing specific and detailed scenarios or using the notion of agent autonomy in this work, one of us has written extensively on the viewpoint that promises represent a sound engineering discipline, somewhat analogous to an "atomic theory", where arbitrary systems can be reduced to independent agents that make promises about their behaviour[6]. This sets about a process of documentation of independent properties that can be helpful in detailing one's understanding of observed phenomena. Indeed, in computing in particular there is a genuine case to be made for viewing a computer as a number of independent electrical components that make quite specific promises to one another.

Given the role promises have in influencing certainty, an exploration of the relationship to causation and fault propagation networks seems fruitful. Similarly the benefits in economics are obvious. Given their value to different parties, promises naturally take on the role of 'strategies' in multi-player economic games and thus provide a basis for formulating strategic games[13]. Evaluating the value of promises is of course a complex matter. A topic for future research would be to consider the existence of transformations between the valuation viewpoints of different agents to see whether there can be simple relationships with well-defined transformations between them (a theory of relativity).

11 Conclusions

Without attempting to suggest applications in any field, we have argued for the usefulness of promises as an independent and practical concept, whether in philosophy, economics or technology. We have compared promises to the more usual idea of obligations and have concluded unequivocally that promises are a simpler theoretical notion and a more practical tool than obligations in the reduction of an agent's uncertainty about the behaviour of other agents.

We show that intentions, promises and commitments can be explained in the absence of an understanding of obligations. Furthermore, although it seems to be a common view that obligations are a more primitive concept than promises, our paper suggests the contrary. Promises need not be viewed merely as proxies for obligations; if promises give rise to obligations, this can in fact be studied purely in an exposition based on promises, intentions and commitments. Indeed more often than not promises are made by agents who would not be able to explain the extent to which their promises might lead to obligations or not, or to what extent such obligations would be more significant than the promises from which they arose.

We expect to return to describe more features of promises in detail in future work.

References

[1] P.S. Atiyah. *Promises, Morals and Law*. Clarendon Press, Oxford, 1981.

[2] J.P.W. Cartwright. An evidentiary theory of promises. *Mind (New Series)*, 93(370):230–248, 1984.

[3] C. Fried. *Contract as promises*. Harvard University Press, 1981.

[4] M. Gilbert. Is an agreement and exchange of promises? *Journal of Philosophy*, 90(12):627–649, 1993.

[5] J.D. Carrillo and M. Dewatripont. Promises, promises. Technical Report 172782000000000058, UCLA Department of Economics, Levines's Bibliography.

[6] Mark Burgess. An approach to understanding policy based on autonomy and voluntary cooperation. In *IFIP/IEEE 16th international workshop on distributed systems operations and management (DSOM), in LNCS 3775*, pages 97–108, 2005.

[7] M. Burgess. Promise you a rose garden. http://research.iu.hio.no/papers/rosegarden.pdf.

[8] Feng Wan and Munindar P. Singh. Commitments and causality for multiagent design. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, 2003.

[9] M. Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, Chichester, 2002.

[10] *The Strategy of Conflict*. Harvard Univesity Press, Cambridge, Mass., 1960.

[11] J. Bergstra, I. Bethke, and M. Burgess. A promise algebra framework for promise theory. Technical report, CoRR abs/0707.0744, 2007.

[12] M. Burgess and S. Fagernes. Laws of systemic organization and collective behaviour in ensembles. In *Proceedings of MACE 2007*, volume 6 of *Multicon Lecture Notes*. Multicon Verlag, 2007.

[13] Promise theory website. http://project.iu.hio.no/promises.php.