

Promise Theory as a Tool for Informaticians

Jan A. Bergstra

janaldertb@gmail.com

Minstroom Research BV, Utrecht, The Netherlands

Submitted: 4 November 2019

Revised: 31 March 2020

Abstract

Promise theory was designed and developed from 2005 onwards by Mark Burgess and his coworkers. It totalises the notion of a promise so that it applies to both animate and inanimate promisers. The focus of promise theory is on applications in informatics and systems design. This paper extends the account of promises by providing more detailed requirements on promises. In particular, the requirement of determinacy, the requirement of decomposition of aggregate agents, and the feature of a promise bias are introduced. The paper further includes an account of threats, as well as of risks, both viewed as an extension of promise theory. It is finally indicated by means of a series of informal examples how and where various kinds of promises and threats may occur in informatics.

1 Introduction

Promise theory was designed by Mark Burgess in a series of papers from 2005 onwards. For a survey of this work as well as some extensions of it I refer to Burgess 2015 [13] and to the more technical, and for that reason unfortunately less readable exposition in Bergstra & Burgess 2012/19 [4]. Bergstra & Burgess 2017 [5] provides an extensive case study for promise theory outside the realm of informatics. Bergstra & Burgess [6] (chapter 11) provides an application of promise theory to the description of informational monies.

It appears that promise theory is useful for understanding systems of animate agents as much as it is for specifying systems consisting of inanimate agents.

Promise theory is based on four assumptions: (i) that a system of cooperating agents can to a significant extent be specified in terms of the promises which each agent is willing and able to make, (ii) that even the simplest entities and objects (e.g. records in a database) can be understood as agents capable of promising and being promised to, (iii) that promising does not create obligations of any kind, unless explicitly stated, and (iv) that successive and cumulative exchanges of promises underlie many fundamental interaction protocols.

According to Mark Burgess these assumptions position promise theory as a useful tool for informaticians. For the use of the term informatician and related terminology I refer to Bergstra 2012 [1]. The objective of this paper is to contribute to Promise Theory by extending it in several ways:

1. To introduce the distinction between structural promises and operational promises, as well as a similar distinction for threats. In the context of promise theory, promises are by default understood as operational promises.
2. To characterise the notion of a promise in more detail than has been done until now in works on Promise Theory. Promise theory is then understood as the theory and application of operational promises.
3. To specify threats as a concept related to but distinct from promises. Promises and threats are quite closely connected themes so that that existing or forthcoming theory of threats may be considered as part of (existing or forthcoming) Promise Theory. Threats were introduced in promise theory in Bergstra 2019 [9], from which this part of the paper has been extracted.
4. To provide some examples of promises and threats as may be found in informatics.

1.1 Defining promises

In Bergstra & Burgess [3] the following definition of a promise is given: a promise is a documented intention. This is at best a provisional definition which in various ways may be compromised. What makes the notion of a promise especially difficult to pin down is its dynamic character, which we have tried to capture below in Paragraph 1.5.

Having defined a promise, it must be acknowledged that intentions, as meant in this definition, are not limited to human agents or to living entities. If a database states that the weight of a car has a certain value, the database intends to make its users believe that said weight has said value. Of course one may claim that it is the owner of the database who has such intentions, rather than the database proper, but we claim that for promise theory it is helpful not to insist on such scrutiny and to think of the database as an entity which might have intentions and which is more or less trusted by other agents and such that the dynamics of the level of trust depends on the degree to which its promises turn out to be kept.

1.2 Promise theory and total systems

A basic assumption of promise theory is that it extends the ability to issue a promise as well as the ability to become aware of a promise beyond humans and beyond animate beings to all entities for which it makes sense to view these as

agents in the context of a description and an analysis of a system. For instance the number zero may promise that it has no predecessor, and zero may promise that applying the inverse function to it produces a value.

The paradigmatic case for total systems is that the predicate “leading to a value upon application of the inverse function” is supposed to reach out to all numbers including zero. In general striving for totalization of systems, in some specific setting, involves a focus on one or more properties which are extended beyond conventional boundaries. After making the transition over such a boundary one may in some cases find mathematics which is similar in spirit but not in form to conventional mathematics. The different approaches for division by zero provide examples of that phenomenon (see Bergstra [2]). For promise theory a second aspect of totality is that making a promise always has some impact, at least in principle. A promise changes the expectations of other agents, and upon assessment of a promise, an agent making that assessment will update its assignment of trust to the agent from which the promise originated. Promise theory assumes that relative trust is a total function on ordered pairs of agents, and that (subjective) expectation is a total function on pairs of agents and events. This function is changing in time, and is influenced by promises which have been made. Trust also regulates the impact which promises have on promisees in terms of their assessment of expectations. And such assessments in turn influence promisee behaviour.

It must be admitted that there is no simple embedding of promises in any form of logic. That lack of a path to clarity seems to correspond with a power of expression, however. Or in other words, those notions from logic which admit relatively unambiguous definitions seem to have relatively limited application at the same time.

Another line of thought which connects promise theory with total systems is the way it supports inclusiveness. Validity, or truth is not the first concern with promises, while the interaction aspect is of prime importance: who said what to whom. When providing a promise theoretic account of a theme there is no need for an author to provide an assessment of the promises which they assign to various actors. In the study of total systems in mathematics, syntax may precede semantics, so that a meaning is sought for an expression the existence of which has been accepted, even if its meaning has not been wholly determined. One may say that, in total systems of mathematics, value may be secondary to form, which seems not to be the case in the conventional interpretation of mathematical text and speech. Promise theory allows a similar understanding of a promise as an entity or a concept the meaning of which emerges from its existence in potentially unforeseen ways, just like the interpretation of $1/0$ in elementary mathematics. Thus $1/0$ is not primarily “the meaning which the expression $1/0$ is supposed to express”, but it is “one’s favourite option for assigning a meaning to the expression $1/0$ given the objectives one has in mind”, with the understanding that these objectives have not been set in stone in an inflexible manner.

Facilitating inclusive accounts of promises is especially relevant in political themes where agents may be in extreme disagreement over the facts, and may

produce widely diverging promises about the state of affairs.

1.3 Promise theory and informal logic

We view promise theory as an approach to informal logic. This idea requires some explanation. Suppose an agent Q contemplates a situation in which assertions A and $\neg A$ are both somehow “given” to Q . What can Q do about this paradoxical situation. Here are some options:

- Q may hold that it cannot possibly be the case that both A and $\neg A$ are valid so that the matter must be investigated in depth so that one of both options may be discarded.

In many cases Q may be unable to find out which of the two “facts” is really true, whatever that means. In such cases this option provides Q little support.

- Alternatively Q may embrace paraconsistency (for a recent survey see Middelburg 2011 [16]), and may claim that the combined assumption of A and $\neg A$, in spite of being uninformative about A , is not destructive for reasoning about other assertions, thereby avoiding the explosive impact of a contradiction in classical logic.

Adopting this view does not bring Q any nearer to the resolution of the paradox, while only its impact is limited. The paraconsistent approach fails if the assessment of a paradoxical assertion is of key importance, and does not focus on a resolution of the resulting dilemma.

- Subjective probability theory suggests that P , taking all available information into account expresses their uncertainty about A by assigning to a certain subjective probability $P(A) = P_Q(A)$. It is plausible that $0 < P(A) < 1$.

Determination of $P_Q(A)$ is unlikely to produce results which admit convincing justification.

- With promise theory as a modelling technique Q may look at the paradox $A \wedge \neg A$ in a dynamic manner. Suppose that P_1 has promised A and that P_2 has promised $\neg A$. Now Q may quantify their trust in agent R as a rational $\tau_Q(R) \in [0, 1]$ where 0 represents no trust and 1 represents full trust, and from these data Q may derive a belief $P_Q(A)$ in A with the rule $P_Q(A) = (1 + \tau_Q(P_1) - \tau_Q(P_2))/2$. Initially Q may set $\tau_Q(P_1) = \tau_Q(P_2) = 1/2$ so that $P_Q(A) = 1/2$. This initial assessment is quite comparable to an approach via subjective probability theory, viz. the determination of so-called prior odds.

At some moment, however, Q may differentiate its trust in P_1 and in P_2 after, say P_1 having kept an important promise A' and P_2 having failed to keep yet another promise A'' with as an effect that $\tau_Q(P_1)$ increases, and that $\tau_Q(P_2)$ decreases, so that in combination $P_Q(A)$ increases. Said

increase occurs as a consequence of information becoming available to Q concerning the keeping or not keeping of entirely different, and logically unrelated, promises, by P_1 and P_2 .

In spite of its complexity at first sight, I believe, the approach via promise theory to the resolution of an apparent paradox has much to offer in cases where an appeal to the facts (the first option) is unfeasible.

1.4 Promise theory versus process theory and temporal logic

Promise theory is unlike various process theories and temporal logics in that it is uncommitted to the type of a promise body A . The following options constitute a non-exhaustive survey of promise bodies A .

- A is an assertion, the truth value of which will not change but which may be hard to determine (e.g. “there is life on Mars”, or “the twin prime conjecture is wrong”, or “the Riemann hypothesis is provable in ZFC”). A may be assessed as unknown until the truth about the matter has been discovered, and the promise is about the eventual outcome of assessments.
- A is an action performed by human agents which takes place at a certain moment of time, the body may specify a time interval. For instance with rationals t, r such that $t < r$, a promise body $A(t, r)$ indicates that A is promised to occur in the time interval (t, r) measured in absolute time, and a promise body $A[t, r]$ indicates that the promiser promises A to occur in the interval $[t, r]$ relative to the moment the promise is made.
- A is an event the causation of which is beyond human control which takes place at a certain moment of time, the body may specify a time interval.
- A may be an assertion which once having become true will stay true. Now $A_{\leq}(t)$ may express that A is promised to become true at (absolute) time t or before, while a promise body $A_{\geq}(t)$ expresses the promise that A will turn true after t . $A_{\leq}[t]$ and $A_{\geq}[t]$ express corresponding promises in relative time.
- A may represent a progression of subsequent (timed) actions and (timed) events of an assertion becoming true or false.

In comparison with current process theories and temporal logics, promise theory is weakly typed with respect to the promise bodies.

1.5 An intuition for promises

In an attempt to characterise intuitively what a promise is, we arrive at the following:

1. By promising A to promisee Q a promiser P initiates a pattern of successive assessments by Q of (i) the subjective probability of the promise being kept, and (ii) of the trust of Q in P , where (iii) this pattern interferes with Q 's own actions which in part may be triggered by the outcomes of said assessments.
2. Similar patterns occur in all agents in scope of a promise.
3. The dynamics of trust is in part determined by communication between agents in scope of a promise via mechanisms of reputation. Such mechanisms are specific for different applications of promise theory.
4. There is no a priori upper limit to the complexity of the process which the issuing of a single promise may create.
5. The side effect of keeping or not keeping a promise on the development in time of the trust of other agents in a promiser is central to the concept.
6. It is not assumed that keeping (not keeping) a promise will increase (decrease) a promisee's trust in a promiser, the dynamics may be the other way around.
7. An obligation on the promiser is created only by the promiser explicitly including the obligation in the promise body. Moreover obligations must be mentioned as well as qualified. Some possible qualification: (i) legal obligation (in which jurisdiction), (ii) moral obligation, (iii) moral obligation in compliance with a specific moral philosophy, (iv) obligation "as a friend", and (v) an obligation which persists for the promiser even if in due time the promisee rejects the offer implicit in the promise, (iv) for each of these options it must be clear whether or not the promiser is bound only under the condition that the promisee has promised to accept the offer made in the promise.
8. In some cases a promise has as a side effect the acquisition (transfer) of authority of the promisee over the promiser's behaviour. This feature is considered central in the theory of promising of Owens 2006 [20], in our promise theory the role of authority transfer is not (yet) considered central.

2 Structural versus operational promises

It is common to speak of the promise of quantum computing and the promise of machine learning and similar promises which are implicit in a societal trend rather than having been explicitly made as a step in the unfolding of an engineering process or of an operational process. I will speak of a structural promise if the term promise is meant to represent a positive outlook for a bundle of potential further developments. Such developments may range from technology to politics, and may be morally judged in different ways. An operational promise

instead is a promise viewed as an action performed by an agent. A more precise description of operational promises is given below. By default I will refer to an operational promise as a promise. Thus instead of the promise of quantum computing I would prefer to speak of the structural promise of quantum computing in case some ambiguity might arise. Structural promises are emergent properties of systems and developments. Structural promises are discovered, observed, or claimed rather than issued or made.

A structural threat is a threat which emerges from a structural development. Sea level rising is a structural threat. Threats as meant in Paragraph 3.1 below are operational threats. By default I will assume that a threat is an operational threat.

2.1 Structural promises versus implicit promises

An implicit promise occurs if an agent acts in such a way that explicitly issuing a promise would be redundant. If a shop sells a new laptop its staff member promises the absence of personal data on the laptop's memory. This promise is so obvious that it is not made explicitly. It could be issued as a promise by the staff member in which case that would be felt as redundant. Implicit promises share with structural promises that these are not operational. When specifying a system it may be useful to describe agent behaviour in part by means of implicit promises. This can be done in a specification by rendering implicit promises explicit.

2.2 Structural threats versus implicit threats

An implicit threat is a threat which, without being issued, is still present and the existence of which can be inferred by an agent. For instance the presence of dogs behind a fence constitutes an implicit promise of problems when trespassing without proper admission. A sign "beware of the dogs" turns the implicit threat into an explicit one, even in the absence of dogs, in which case the threat is a deception.

2.3 Promise and threat analysis

Once a structural promise has been discovered it may be turned into an operational promise by any agent who knows about it. The main consequence of a promise having been turned (or rather cloned or copied) into an operational promise is that promiser and promisee are now specified and side effects on trust in the promises become unavoidable.

Structural promises are objective in the sense that scientific research may reveal or discard structural promises. Structural promises may be hypothesised and may be the subject of extended disputes by experts coming from different backgrounds and using approaches.

2.4 Risk versus threat

Closely related to the concept of threat is the concept of risk. Many definitions or risk are possible:

A risk is the possible consequence of a (structural or explicit) threat (of type [+], see below) coming true. The threat may either have gone unnoticed (but may yet be revealed via risk analysis) or it may have been revealed. The risk is yet unquantified in terms of subjective probability of occurrence and in terms of the magnitude (impact) of damage or consequences. Quantifying the risk involves the integration of subjective probability of occurrence and estimation of impact.

A large volume of theory involving threats has been developed within the area of nuclear deterrence, see for instance Powell 1985 [21]. Those uses and applications of threats are not based on an underlying account of promises, however.

2.5 Promise versus lie

If A promises to B that A has made a transfer of an amount a of money to B , then this utterance may qualify as a lie, at least when taking the first definition of a lie in Mahon 2016 [18] as a criterion for lying. Specifying the relation between promises and lies in more detail, however, requires that neighbouring notions like fact, alternative fact, news, fake news, deception, and opinion are taken into account as well. These considerations have not yet been detailed in a satisfactory manner in the context of promise theory and will require further research.

3 A characterisation of (operational) promises

In order for promises to play the role of operational promises the notion of a promise is supposed to be characterised and restricted in the following manner:

Promising is an action. Issuing a promise is an action performed by an agent.
(Agents may be animate or inanimate.)

Documented intention. Issuing a promise amounts to the production and distribution of a documented intention. A promiser may in fact not intend to keep a promise, in which case the promise is a deception (on top of being a promise).

Promiser, promisee, and scope. Each promise has a promiser, that is an agent which is issuing the promise, a promisee, that is an agent to which the promise is primarily directed, and a scope consisting of zero or more agents who notify the promise being issued. Agents in scope of the promise maintain a record about it which is used for updating its status when following a trajectory through its life-cycle (see below).

Promise body. The body specifies the content of what is promised.

Promise qualifier. Qualifiers provide a typing of promise bodies with offering (denoted with [+]) and accepting (denoted with [-]) as the main types. Promise qualifiers are surveyed below in Paragraph 4.1.

(Optional) condition(s). A promise may be conditional, if so it has one or more conditions, where a list of conditions is understood as a conjunction of these. A condition may be any action or state of affairs (including the keeping of another promise). Upon a condition being satisfied and that fact becoming know to an agent maintaining a record of that promise, the promise (or rather its record) is transformed into a promise with fewer conditions, and upon it being refuted the promise is rendered vacuous and terminates for that reason. If each of the conditions is assessed as being satisfied then the conditional promise is transformed into an unconditional promise.

Promise life-cycle. A promise moves through a life-cycle. More specifically each agent in scope of the promise maintains a copy of it which moves through its own life-cycle. A (copy of a) promise may terminate by being kept, being (explicitly) not kept, being withdrawn, and by being forgotten. Agents in scope of the promise maintain a record of it which serves to store the state of the promise in its envisaged life-cycle.

Deontic neutrality. Issuing a promise does not by itself create an obligation, be it a moral, a legal, or a contractual obligation. If an obligation of some kind and with some way of enforcement must be created all of that should be included in the body explicitly or be included in one or more additional promises.

Side effects on trust and expectation. Issuing a promise primarily has side effects of the following kind:

1. the promisee as well as other agents in scope adopt expectation values of the promised state of affairs or activity as well as of related states of affairs and activities,
2. continuously through the life-cycle of the promise (and finally upon termination of its life-cycle) various agents in scope of the promise update their trust in the promiser depending on their assessment of the likelihood that the promise is kept,
3. a promiser's failure to withdraw a promise, whose likelihood of being kept has severely degraded, may notice or experience the consequences of a negative side-effect of this failure on the trust that other agents have in the promiser.

Subjective necessity caused by risk of losing trust. A promiser or other agents in scope of the promise, may fear the loss of trust resulting from not keeping a promise so much that a perception of necessity (that is

the awareness of a very high priority), concerning keeping the promise, results. This sense of necessity is subjective and it is to be distinguished from obligations of any form.

(Optional) promise bias. Optionally a promise may be understood best in the context of a bias held by the promiser concerning a certain point of view or objective. For instance the enthusiastic organiser of a trip may promise potential guests that the weather will be perfect. Of course perfect weather may be useful for alternative activities as well.

Determinacy. The meaning of a promise body is supposed to be reasonably clear. Of course this depends on the context but the assumption is that the promisee is satisfied with its ability to figure out what is meant, rather than that the promise triggers extended guesswork about that.

For instance if A promises B to pay some amount at some future instant of time, this body will not qualify as being sufficiently determinate in the majority of contexts. The problem (lack of determinacy) with this particular promise body is the lack of information on vital parameters: how much must be paid, in which currency will it be transferred, by what means will it be done, and for what purpose.

Promise assessment. Assessing whether or not, and to what extent, a promise is kept, is a matter for each agent in scope of the promise. The same promise may be assessed repeatedly by the same agent. The result depends on the past history for promiser, promisee, and other relevant agents. Moreover the assessment, for the same promise may vary between different agents, even when performed simultaneously.

If a conditional promise depends on the assessment of the fate of another promise being kept or not kept, then different agents may at the same time maintain diverging views on what happened to the original promise serving as a condition. If a promiser has made a promise under the condition that another promise is kept, the promiser may repeatedly perform an assessment of whether or not the condition (viewed as a promise) is kept.

If A promises B to take B by car to the station assuming that B is at A 's place at time t or before, and B accepts this promise and promises to be at A 's place at 5 minutes before t then in case B does not show up A is likely to make repeated assessments on whether or not B keeps its promise, and say at quarter past t , A may conclude that B failed to keep its promise to be in time so that A can forget about its own promise which has now expired.

(Optional) label. When writing about promises it is sometimes useful to provide a promise with a label that serves as a name during subsequent referencing. Names are invariant during the life-cycle and for each application of promise theory names are supposed to be unique.

Decomposition of aggregate agents as promisers. An aggregate of agents may also be considered an agent (for more information see the compound agents as put forward in Burgess 2014 [10]). If an aggregate agent issues a promise then, if one of those agents fails to keep its part of that promise, such failure is only moderately held against the other agents in the aggregate. In other words a promise issued by a group amounts to little more than (i.e. decomposes into) a combination of promises of its constituent agents. If an agent in an aggregate agent wishes to promise that one or more other agents in the same aggregate will indeed keep their part of the promise such promises must be included explicitly in the promise body, or in the form of additional promises (to be qualified as documented expectations).

This list of criteria serves as a definition of the concept of a promise. At the same time it deviates from a conventional understanding in at least these aspects: deontic neutrality is commonly denied for a promise by assuming that issuing a promise creates an obligation, determinateness is an uncommon criterion, and decomposition of aggregate promisers is not taken into account. Determinateness has been included as a requirement in order to be able to better differentiate promises from threats.

Defining promises by means of the criteria just listed deviates from the colloquial use of the term. Gerring 1999 [14] discusses in detail how concepts may be formed in the case of social sciences. I will assume that requirements engineering and systems design to constitute subdisciplines of informatics which may host the development and use of promise theory. Moreover, requirements engineering and systems design may be viewed as a part of informatics which belongs to the social sciences as much as to engineering. These subdisciplines of informatics may be considered part of social sciences because of the pivotal role played by human engineers primarily working with the tools of natural language. Now these remarks are made to justify contemplating the concept of a promise from the perspective of concept formation in social sciences. Gerring describes that deviations of conventional meaning need not necessarily lead to the use of different terms or the formation of neologisms, and I will assume that this guidance applies in the case of promises. Further Gerring suggests that a good concept can be identified by virtue of eight qualities of it: familiarity, resonance, parsimony, coherence, differentiation, depth, theoretical utility, and field utility. The given definition of promises provides adequate differentiation from related concepts such as: deception, threat, obligation, and imposition. This differentiation goes at cost of parsimony understood as the brevity of a term or phrase as well as of its definition. Gerring acknowledges that defining concepts is a matter of trade-off. Whether or not promise theory conceptualises promises in such a manner that a good score on Gerring's criteria is obtained remains to be seen. Finding an answer to that question may will take additional work, both theoretical and by way of case studies.

3.1 Promise versus threat

Instead of promiser and promisee I will speak of threat source and threat target. A threat is like a conditional promise with these three adaptations:

Negative value to threat target. As an additional requirement: the event of the threat being kept is supposed to be considered a disadvantage for the promisee (the target of the threat),

Uncertainty regarding source intention. The requirement on determinacy is weakened: the promiser may indicate uncertainty about its intention to keep the promise, or may indicate a lack of information about the promised state of affairs, that is determinacy is not required for a threat, and

Quantified subjective probability of keeping for source. Instead of determinacy the quantification of subjective conditional probability (of the body of the threat coming true upon the condition being met) is required. Such quantification may be provided with rational numbers but also with labels for probability ranges. (In the absence of any quantification the threat amounts to an assertion of risk). The subjective probability must be known for the threat source, it may be unknown for the receiver of the threat (the threat target).

There is no assumption that a promise is not a threat. Indeed A conditional threat with a determinate body qualifies as a promise as well. A threat is deceptive if the subjective conditional probability of it being kept (upon the conditions having been met) is low.

Deontic neutrality exists for threats as well as for promises because threats are not considered as binding while primarily serving as a warning or as advice. By requiring deontic neutrality for promises, threats and promises are made symmetric and threats become a special case of a generalisation of conditional promises.

The keeping of a threat may or may not be detrimental to other agents in its scope. For instance (Authorities) *A* may promise citizen *B* that if he has another fight with his neighbour *C*, he (*B* and his family) will be removed from their home. Other neighbours may be in scope of the threat. Now it may be the case that *C* has regular conflicts with several other neighbours as well and these may view removal of *B* upon a further conflict with *C* as unfair and as a weakening of their own position towards *C*. If, however, neighbours in scope of the promise have no conflicts with *C*, they may side with *C* and may applaud the promise having been made.

3.2 Motivating deontic neutrality

The reasons for requiring deontic neutrality are these:

- Agents are supposed to be autonomous as a matter of principle and agents must not violate the autonomy of other agents by issuing a promise. As

a consequence in the presence of say two conflicting promises, it is up to the promiser to determine how to proceed, in particular not to keep at least one of both promises. Stated differently, due to deontic neutrality, there may only arise trade offs between promises that cannot be kept at the same time.

- An agent cannot solve the problem posed by conflicting obligations, not even if the agent has created these obligations all by itself. If obligations are created, for instance as a side product of issuing promises, the origination of conflicting obligations can not be avoided in a plausible manner. As a consequence by allowing promises to create obligations promises become less useful as a tool for distributed self-organisation for agent communities.
- Promise theory is primarily meant as a tool for requirements capture as well as for the design of cooperating artificial agents, and for that reason there is a need to have promise theory available without any mention of ethics or morality.

However, by explicitly promising to accept a moral obligation, a legal obligation, or a contractual obligation, various forms of obligation can be introduced in a system specification involving promises in such a manner that by default promises do not create obligations (that is unless the promise body explicitly features obligation creating constituents).

As a hypothesis underlying this work it is assumed that for use in human context the assumption of deontic neutrality is just as useful as for use in the context of communities of artificial agents.

4 Qualifiers for promises and for threats

Qualifiers provide additional information on the kind of promise or threat. Promise qualifiers support the understanding of promises in cases where conventional intuitions may prove defective. In particular a deviation from common understanding is that a promise may deal with a state of affairs not requiring any forthcoming activity and this idea gives rise to a promise qualifier. For instance a data scientist may promise a manager that client data have been collected with adequate client approval. Such a promise is qualified as a documented expectation, in this case the expectation of what an inquiry about recent data collection methods would reveal.

4.1 Promise qualifiers

A promise qualifier informs about two aspects of a promise: (i) is the promise being issued in response to another promise and (ii) is the promiser's forthcoming behaviour a causal factor for keeping the promise.

In [4] the following notation is used for promises: Each promise has the following structure:

$$\text{Promiser Agent} \xrightarrow[\text{Scope}]{\text{'body'}} \text{Promisee Agent} \quad (1)$$

where the body of the promise represents an explication of the intended outcome, and the scope is a list of agents who are privy to the promise between the promiser and promisee.

Documented intention. + ; promiser claims a causal role in bringing about the expected state of affairs.

These are the promises of the first kind in [4]. With promiser A , promisee C , body B , and scope S a symbolic notation for this promise is:

$$A \xrightarrow[S]{+B} B \quad (2)$$

Acceptance of documented intention. - ; promiser accepts the promise (+ type) made to it (as a promisee) by an agent (in the role of an original promiser) and renders this acceptance as a promise made to the original promiser.

Acceptance is also possible for a threat: the promiser acknowledges and thereby accepts the existence of a threat as issued by the promisee. In the formalism of [4]:

$$A \xrightarrow[S]{-B} B \quad (3)$$

Rejection of documented intention. -! ; promiser indicates unwillingness to accept the promise (+ type) made to it (as a promisee) and renders this rejection as a promise made to the original promiser.

$$A \xrightarrow[S]{-!B} B \quad (4)$$

Documented expectation. [+]; promiser claims no causal role in bringing about the expected state of affairs (expectational promise).

These promises correspond to promises of the fourth kind in [4] provided one accepts the existence of an agent which may bring about the expected state of affairs. A documented expectation promise results from a promise of the fourth kind by abstracting from the latter agent. Documented expectations are weaker than documented intentions because the promiser outsources any responsibility for keeping the promise.

$$A \xrightarrow[S]{[+]B} B \quad (5)$$

Agreement with documented expectation. [-] ; promiser promises agreement with an expectational promise to the promiser of the latter (acceptance of expectational promise.)

$$\mathbf{A} \xrightarrow[\mathbf{S}]{[-] \mathbf{B}} \mathbf{B} \quad (6)$$

Rejection of documented expectation. [-!] ; promiser indicates disagreement with original promiser's expectation cast as a promise to the latter (rejection of expectational promise.)

$$\mathbf{A} \xrightarrow[\mathbf{S}]{[-!] \mathbf{B}} \mathbf{B} \quad (7)$$

Promise qualifiers are alternatively referred to as promise types.

4.2 Threat qualifiers

A threat qualifier informs about two aspects of a threat: (i) is the threat being issued in response to another promise and (ii) is the threat source's forthcoming behaviour a causal factor for keeping the threat.

Formal notations for threats with threat qualifiers are the same as for promises with qualifiers.

Documented intention. + ; threat source claims a causal role in bringing about the expected state of affairs.

Rejection of a documented intention (promise). - ; threat source rejects a promise (+ type) made to it (as a promisee) by an agent (in the role of an original promiser) and returns a threat to the promiser conditional upon them keeping the promise (as intended by the source).

Documented expectation. [+] ; threat source claims no causal role in bringing about the expected state of affairs (expectational promise).

Rejection of documented expectation (promise). [-!] ; threat source indicates disagreement with original promiser's expectation cast as a threat to the latter (rejection of expectational promise.)

Threat qualifiers are alternatively referred to as threat types.

5 Promises and threats in informatics

A useful way to classify promises is as the artifice of promiser and promisee. The applications that Mark Burgess had in mind when developing promise theory each feature artificial agents as promisers and as promisees and make little use of additional agents being in scope of a promise, besides the promisee who is always in scope. Taking into account human agents gives rise to a further classification of promises and threats. In this Section I will work out this classification in some detail and I will provide suggestions for examples in case these exist.

5.1 Four types of promises

Promises and threats may be categorised according to the kind of promiser and promisee (threat source and threat target). Four types of promise result.

Artificial agent to artificial agent. The most ubiquitous case, e.g. a website host promising a bot to deliver a web page in return for a registration. The bot accepting that promise by engaging in the requested registration.

A virus contained in a document provides a promise (to render a document in an orderly manner) which is at fault so that the promise is a deception.

Artificial agent to human agent. This case has come to prominence with ransomware. An infection with ransomware of a host may be considered as the confrontation with an artificial agent which, after entering a system and encrypting important data found on that system makes two promises to the users of the host: (i) upon receiving a specified payment before a given deadline the data will be decrypted (or a key for doing so will be provided), and (ii) in the absence of the required payment the decryption cannot be undone anymore.

For a human agent the question is whether or not to accept the first promise by making the required payment. This is hugely a matter of trust: making the payment is futile if there is no expectation that the data will be decrypted.

A common case of an artificial to human promise is found when an app promises access upon registration conditional on confirmation of the intention to access via another communication channel (SMS, WhatsApp, email etc.).

Human agent to artificial agent. Technically such promises take the form of artificial agent to artificial agent where an artificial agent acts as the proxy for a human agent. However, in case for instance a site asks for filling in a form before offering a specific service, a human user may (in some designs for the site) accept that conditional promise by a single click on a button thereby promising to fill in the form.

A person may, on request, by means of a single click promise to take an interest in regularly being sent emails with information about a specific topic by the promise.

When a site claims that only authorised persons are admitted to it, by proceeding with logging in, a human client promises to be in possession of adequate authorisation. (If that is false the promiser cannot keep the promise, and if the human client knows that the authorisation is absent the promise is in fact a deception.)

Human agent to human agent. Such promises are central to any branch of business. Some examples:

- The promise made by a service provider to a client to regularly upgrade software in order to comply with regulations regarding data protection and data integrity.
- The promise made by a software provider that unless an upgrade is installed before some deadline the client will be in breach of a national regulation on data protection.
- The promise made by a hardware vendor that a laptop which he is selling will still work well and without loss of data after having been submersed in salt water for at most 24 hours.
- Bergstra & Burgess 2019 [7, 8] reports an extensive case study on potential software problems with the MCAS algorithm of the Boeing 737 Max 800 aircraft (as present in the version operational during 2017, 2018 and the first months of 2019).

Promises in a hybrid context, that is involving either a human promise and an artificial promisee, or conversely, are very likely to come to prominence once hybrid networks of combatants involving human fighters as well as intelligent autonomous weapons become a reality (see Nørgaard 2016 [19]).

5.2 Four types of threats

Just as with promises I will distinguish four types of threats.

Artificial agent to artificial agent. I am not aware of an example of this kind of threat, though if ransomware attacks were to be handled automatically such attacks involve threats from machine to machine.

Artificial agent to human agent. The operating system S may issue a warning to its client C : if you do not upgrade your operating system at your earliest convenience you run a high risk of being infected with newly discovered virus V very soon.

Human agent to artificial agent. I am not aware of the existence of this kind of threat. However, by issuing intentionally a false fire alarm in a system where the alarm is initially handled by artificial agents a person might be considered to be in fact threatening a machine.

Human agent to human agent. An example: A software provider A says to client B who bought custom made system S from A : if you do not pay the increased fee for the service contract extension we may (a few months from now) refuse to do any maintenance work for S although we have serviced that system for you until now.

Another example: a software provider states that unless an upgrade is installed before some deadline the client will be in breach of a national regulation on data protection and may be caught by authorities who are in the process of making up their mind regarding carrying out random inspections in such matters of regulation.

5.3 Further distinctions of promises and threats

The scope of a promise or threat may consist of both human agents, say via cookies embedded in some application, and artificial agents. Here are again three options for all kinds of promises and threats:

Only human agents in scope. This case matters with promises issued in informal language between humans. In a political setting scoping is highly important.

Human agents in scope as well as artificial agents in scope. I am currently unaware of example of this configuration of agents in a network involving promising and threatening, but that may just be a matter of time. Sooner or later all observations accessible to humans will also be monitored and processed by machines.

Only artificial agents in scope. This is very much the original motivation of promise theory, where promises were used as a means to specify the intended behaviour of the configuration management tool CFEngine. For a discussion of CFEngine in the context of promise theory I refer to [4].

6 Concluding remarks

Promise theory has been extended with a discussion of threats and risks, leading to various classifications, and with indications of examples for various options within informatics. The assumption of deontic neutrality for promises is the key factor that makes promise theory available as a tool in informatics. This controversial observation, turned into an axiom of promise theory, underlies the original proposals made by Mark Burgess. For threats (and risks) assuming deontic neutrality is uncontroversial which adds to the credibility of including an account of threats and of risk in promise theory. For other observations on the use of promises in informatics and artificial intelligence I refer to [11] and [12].

At present the scope of promise theory is hard to assess, and our primary focus is on the use of promise theory for conceptually fairly simple examples, e.g. in robotics, but ultimately promise bundles may turn out to be useful carriers for larger and less well-delineated complexes including full scientific theories and research paradigms.

References

- [1] Jan A. Bergstra. Informatology: combining Computer Science, Data Science, and Fiction Science. <https://arxiv.org/abs/1210.6636>, (2012).
- [2] J.A. Bergstra. Division by zero, a survey of options. *Transmathematica*, ISSN 2632-9212, (published 2019-06-25 2019), DOI: <https://doi.org/10.36285/tm.v0i0.17>, (2019).

- [3] Jan A. Bergstra and Mark Burgess. *A static theory of promises*. <https://arxiv.org/abs/0810.3294v5>, (2014).
- [4] Jan A. Bergstra and Mark Burgess. *Promise Theory: Principles and Applications*. χt Axis Press. ISBN9781495437779 (2014), second edition (2019).
- [5] Jan A. Bergstra and Mark Burgess. *Promise Theory: Case Study on the 2016 Brexit Vote*. χt Axis Press. Kindle version: ASIN B074V9BW1C, Paperback: ISBN-13 987-1974545339 (2017).
- [6] Jan A. Bergstra and Mark Burgess. *Money, Ownership, and Agency, as an Application of Promise Theory*. χt Axis Press. ISBN9781696588379 (2019).
- [7] Jan A. Bergstra and Mark Burgess. A promise theoretic account of the Boeing 737 MCAS affair. <https://arxiv.org/abs/2001.01543v1>, (2019).
- [8] Jan A. Bergstra and Mark Burgess. Candidate software process flaws for the Boeing 737 MCAS algorithm and risks for a proposed update. <https://arxiv.org/abs/2001.05690v1>, (2020).
- [9] Jan A. Bergstra. *Promises and Threats by Asymmetric Nuclear-Weapon States*. χt Axis Press. ISBN9781673128215 (2019).
- [10] Mark Burgess. Spacetimes with Semantics. <https://arxiv.org/abs/1411.5563>, (2014).
- [11] Mark Burgess. Spacetimes with Semantics II. <https://arxiv.org/abs/1505.01716>, (2015).
- [12] Mark Burgess. Spacetimes with Semantics III. <https://arxiv.org/abs/1608.02193>, (2016).
- [13] Mark Burgess. *Thinking in Promises: Designing Systems for Cooperation*. O'Reilly Media (2015).
- [14] John Gerring. What Makes a Concept Good? A Criterial Framework for Understanding Concept Formation in the Social Sciences. *Polity* 31 (3) 357–393, (1999).
- [15] J. Andrew Jensen. Nuclear Weapons Proliferation Theory and the Case of Iran. *Hinckley Journal of Politics* 9, 29–41, (2008).
- [16] C. A. Middelburg. A survey of paraconsistent logics. <https://arxiv.org/abs/1103.4324>, (2011).
- [17] Bruno Latour. *Reassembling the social: An introduction to actor-network-theory*. Oxford university press, (2005).
- [18] James Edwin Mahon. The definition of lying and deception. The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/win2016/entries/lying-definition/> (2016).

- [19] Katrine Nørgaard. Autonomous Weapon Systems and Risk Management in Hybrid Networks. In: In “The United Nations Convention on Certain Conventional Weapons”. *Royal Danish Defense College* (2016).
- [20] David Owens. A simple theory of promising. *Philosophical Review* 115 (1) 51–77, (2006).
- [21] Robert Powell. The Theoretical Foundations of Strategic Nuclear Deterrence. *Political Science Quarterly* 100 (1) 75–96, (1985).