UNIVERSITY OF CAPE TOWN
DEPARTMENT OF COMPUTER SCIENCE

CS/IT Honours
Final Paper 2019

Title: Non-Classical Reasoning : On the Relation Between Cognitive Reasoning and Belief Revision

Author: Claire Denny

Project Abbreviation: CDR

Supervisor(s): Professor Thomas Meyer

Category	*Min*	*Max*	Chosen
Requirement Analysis and Design	*0*	*20*	0
Theoretical Analysis	*0*	*25*	5
Experiment Design and Execution	*0*	*20*	20
System Development and Implementation	*0*	*20*	0
Results, Findings and Conclusion	*10*	*20*	20
Aim Formulation and Background Work	*10*	*15*	15
Quality of Paper Writing and Presentation	*10*		10
Quality of Deliverables	*10*		10
Overall General Project Evaluation (*this section allowed only with motivation letter from supervisor*)	*0*	*10*	0
Total marks		**80**	

Non-Classical Reasoning : On the Relation Between Cognitive Reasoning and Belief Revision

Claire Denny
dnncla004@myuct.ac.za
University of Cape Town

ABSTRACT

In reality, decision-making is commonly performed without complete information and certainty, exceptions exist and knowledge is not static. Classical reasoning, being monotonic i.e. deductive, is too inflexible to facilitate the style of reasoning this necessitates. Non-classical forms of reasoning model the non-monotonic nature of human reasoning. Belief Revision is such a form of reasoning. It can be described in two ways: using semantics and using postulates. We give a description using postulates. Using this, we explore cognitive reasoning with respect to Belief Revision by means of a survey. Findings suggest that Success, Closure and Vacuity are employed in cognitive reasoning. It is also found that whilst normative reasoning styles are employed in human reasoning, descriptive reasoning styles are more prevalent.

CCS CONCEPTS

• **Theory of computation → Logic**; • **Computing methodologies → Nonmonotonic, default reasoning and belief revision**; *Reasoning about belief and knowledge*;

KEYWORDS

propositional logic, defeasible reasoning, cognitive defeasible reasoning, belief revision, belief update, non-monotonicity, propositional logic, survey, Mechanical Turk, Google Forms

1 INTRODUCTION

In their everyday lives, humans are faced with incomplete knowledge but still must act [37]. As a result of their incomplete explicit knowledge, humans rely on background knowledge, heuristics and their logic's capacity to be flexible in what they believe about the world, allowing themselves to withdraw previously made conclusions, if necessary, and make new ones, given new evidence [29]. This flexibility in their reasoning classifies it as non-monotonic. Makinson [29] formally defines non-monotonicity. Consider a medical practitioner or a mechanic - people whose professions require diagnoses of problems such that the appropriate action can be performed. In contexts where an action is urgent, the time required to reach deductive certainty may be too long [29], necessitating non-monotonic logic such as default logic. Lehmann and Magidor [26] acknowledge that humans are remarkably good at making and correcting inferences using their knowledge bases.

Noting the importance of the property of flexibility in intelligence [36], which humans demonstrate daily, progress in artificial intelligence hinges on non-monotonic reasoning and its interactions with other intelligent activities being further researched [37].

We want to examine the closeness between how humans and artificially-intelligent computers reason. The link between human reasoning and Belief Revision has not been investigated extensively - this research aims to reduce that gap. Moreover, the results of the project will be another step towards the greater goal of understanding human cognition.

Using classical propositional logic and its notation as a starting point, this paper considers a type of non-classical reasoning within Belief Change, called Belief Revision. It reviews past work in the field and explores the relation between cognitive reasoning and Belief Revision through the conducting of an experiment.

Two questions were explored through this experiment:

1. Which of the properties of Belief Revision, as formalised by Alchourrón, Gärdenfors and Makinson (AGM) [1], are employed by humans when reasoning?

The flexibility that characterises belief revision as non-classical is a key feature modelled in cognitive reasoning. Alchourrón, Gärdenfors and Makinson [1] propose eight postulates to further characterise the process of belief revision, where six of the postulates are basic and two are supplementary. For each postulate, we are interested in whether it is obeyed in the way humans reason, and if so, to what extent. This is investigated by means of a survey comprising questions tailored to test for each property's use.

2. Which reasoning style is more prevalent among the participants: normative or descriptive?

In the defeasible reasoning community, there is an emerging research project to test whether the normative properties of formal systems of defeasible reasoning are appropriate for modelling human reasoning. Besold and Uckelman [3] argue that the formulations of logic in Artificial Intelligence are normative, but built on a descriptive notion of human reasoning. We are interested in whether the formalisation of Belief Revision, as propounded by Alchourrón, Gärdenfors and Makinson [1], compares normatively or descriptively to cognitive reasoning. This is tested empirically, with the aim of exploring the philosophical link between Belief Revision and cognitive reasoning.

2 BACKGROUND

There is an interest in how non-classical forms of reasoning relate to human reasoning, given that human reasoning prompted the existence of research in this area. Our focus is on belief revision. Defeasible reasoning and belief update, two other forms of non-classical reasoning, are included here for context, as related work. All three extend from classical propositional logic.

The literature proposes systems that perform non-monotonic inferences [23], and evaluates these systems according to formalised characteristics noted as desirable [1, 18, 23]. Reference is made to non-monotonic systems modelling characteristics of the way in which humans reason in the face of incomplete information and uncertainty [38, 40]. Whether such formalisations match human reasoning is an open question.

2.1 Classical Propositional Logic

Propositional logic uses a formal language based on an alphabet of propositional variables [40]. Ragni [40] tells us that propositions are statements with a Boolean value (*true* or *false*), the simplest of which are referred to as atoms, as they are indivisible. Propositional logic focuses on the ways in which statements can be combined or modified and the properties and relationships that arise from doing so.

There is a property called monotonicity that all classical propositional operators must satisfy [23]. Makinson [29] defines monotonocity as the principle that, given a set A, of propositions, if β follows from A, then β follows from any set B, where B $\supseteq$ A.

2.1.1 Notation and terminology. Reference will be made to interpretations, worlds or states of the world. These terms are equivalent, and refer to assignments of truth values for the relevant propositional alphabet [26]. A set of statements explicitly known about the world, known to the reasoner or agent, is called a knowledge base. The symbol $\models$ represents entailment or logical consequence. A model of α refers to a state of a world, ψ, where α is true, which is to say $\psi \models \alpha$ [40]. The notation m(α) can also be written as mod(α) and means the models of α, which is to say that $mod(\alpha) = \{\psi | \psi \models \alpha\}$ [40]. Saying that knowledge base K entails some statement α i.e. $K \models \alpha$, we have intuitively that m(K)$\subseteq$m(α), as if K contains statements in addition to α, the set of models of K would be more constrained than the set of all worlds where it is just α that must hold. In classical logic, there are also the following connectives with which all truth-functional connectives can be expressed [29]: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction). Material implication (if) is denoted as $\Rightarrow$ and equivalence (iff) is denoted as $\Leftrightarrow$. Classical consequence is denoted $\vdash$ and is considered such if it is consistent with what is known about the world.

The logic systems reviewed in later sections will assume an underlying logic that includes classical propositional logic.

2.2 Belief Revision

Belief Revision is a form of belief change [20]. Belief change involves a belief base and a belief set [10]. Explicit knowledge the agent has about the world resides in the base, and inferences or knowledge derived from that in the base resides in the belief set. We have a classical knowledge base assumed to be correct and assume a propositional language with a Tarskian consequence relation $C_n(\cdot)$ [4].

In the event of new information α, inconsistent with the knowledge base K, the situation must be handled such that consistency is restored, or preserved. In belief revision, conflicting information indicates flawed prior knowledge on the part of the agent, forcing the retraction of conclusions drawn from it [20, 31]. The defeasible reasoning approach would be to flag the clash as defeasible, weakening the propositions previously in K that conflict; the *belief revision* approach, within belief change, would be to revise K by invalidating worlds which are sufficiently far from α [18], which is to say that the beliefs are modified such that there is consistency. This means that information is taken into account by selecting the models of the new information closest to the models of the base, where a model of information μ is a state of the world in which μ is true [20].

Alchourrón, Gärdenfors and Makinson (AGM) concern themselves with three operations on the knowledge base, which is to say, three types of theory changes: expansion, contraction and revision [1]. Expansion is performed where new, consistent information is added to theory or knowledge base K and the expanded set is closed under entailment [1, 4]. Contraction is performed where a proposition *alpha* previously in a theory or knowledge base K is rejected [1]. The operation yields a K' that does not entail α [4] - we say the contraction of K by α. It must also, however, be closed under entailment, so other propositions than α may need to be rejected along with α [1]. Revision, as discussed earlier, is performed on K in the case of proposition α such that the resulting K' is both consistent and closed under entailment [1, 4]. It can be seen as the composition of two sub-operations: expansion of the contraction of K by $\neg\alpha$, by α. Using notation introduced by Casini et al. [4], this can be written as $K^*_\alpha = (K^-_{\neg\alpha})^+_\alpha$, which Hansson [13] terms the *Levi identity*. In the AGM model [1], the operator $*$ is that of *partial meet* revision.

2.2.1 Notation and terminology. $C_n(\cdot)$ is an operation that takes a set of propositions K and returns a set of propositions K' [1]. Notation-wise, we can write $\beta \in C_n(A)$ as $A \vdash \beta$. This consequence relation includes classical tautology, is compact and satisfies *introduction of disjunction in the premises* [1]. A *theory* refers to a set of propositions closed under C_n. A set of propositions or beliefs, A, is *consistent* modulo $C_n \Leftrightarrow \beta \wedge \neg\beta \notin C_n(A)$ for any proposition β.

2.2.2 Properties. The following are the properties or postulates of Belief Revision as defined by [1]. Any function $*$ that satisfies them is considered an AGM revision function. Properties 1-6 are the basic AGM properties specifically for belief revision, and properties 7-8 are supplementary AGM properties [4, 38].

1. Closure : K $* \alpha = C_n$(K $* \alpha$)

This implies logical omniscience on the part of the ideal agent or reasoner, including after revision of their belief set [38].

2. Success : K $* \alpha \models \alpha$

This expresses that the new information should always be part of the new belief set [38]. Peppas [38] also considers ways to relax this property, given that it places substantial trust in the reliability of α.

3. Inclusion : K $* \alpha \subseteq C_n$(K $\vee \{\alpha\}$)

4. Vacuity : If $\neg\alpha \notin$ K then C_n(K $\vee \{\alpha\}) \subseteq$ K $* \alpha$

Properties 3 and 4 are motivated by the principle of minimum change [38]. Together, they express that in the case of information α, consistent with belief set or knowledge base K, belief revision involves performing expansion on K by α i.e. none of the original beliefs need to be withdrawn.

5. Consistency : K $* \alpha = C_n(\alpha \wedge \neg\alpha)$ only if $\models \neg\alpha$

This expresses that the agent should prioritise consistency, where the only acceptable case of not doing so is if the new information, α, is inherently inconsistent - in which case, *success* overrules *consistency* [38].

6. Extensionality : If $\alpha \equiv \phi$ then K $* \alpha =$ K $* \phi$

This is also known as the *irrelevance of syntax* postulate. It effectively expresses that the content i.e. the belief represented, and not the syntax, affects the revision process, in that logically equivalent sentences or beliefs will cause logically equivalent changes to the

belief set [38]. This property would not hold without the notion of *epistemic entrenchment* (degree of resistance to change [38]) or Katsuno and Mendelzon's treatment of integrity constraints [18].
7. Superexpansion : $K * (\alpha \wedge \phi) \subseteq C_n(K * \alpha \vee \{\phi\})$
8. Subexpansion : If $\neg\phi \notin K *$ then $C_n(K * \alpha \vee \{\phi\}) \subseteq K * (\alpha \wedge \phi)$
Properties 7 and 8 are motivated by the principle of minimal change [38]. Together, they express that for two propositions α and ϕ, if in revising belief set K by α one obtains belief set K' consistent with ϕ, then to obtain the effect of revising K with $\alpha \wedge \phi$, simply perform expansion on K' with ϕ. In short, $K * (\alpha \wedge \phi) = (K * \alpha) + \phi$.

2.2.3 Discussion. A belief state can be modelled by a belief set i.e. by a set K of sentences that is closed under logical consequences [10]. There is an argument for the need of a *belief base* B_K for a belief set K [10], where the base contains the explicit beliefs or beliefs of independent standing and the belief set K comprises $C_n(B_K)$. In this way, a distinction is made between basic and derived beliefs. The idea is that revisions are performed on the finite belief base, as opposed to the infinite belief set [10].

Rational consequence relations and revision operators are linked - the rules on the former can be interpreted in terms of the latter [19]. Moreover, belief change has connections to non-monotonic inference [28]. Makinson and Gärdenfors [30] study these on a syntactic level. Casini et al. [4] take this idea and previous results from a paper by Casini and Meyer, and explore integrating belief change and non-monotonic inference. They do this by looking at belief change for a preferential non-monotonic framework. They are not the first to study belief revision in a conditional framework [4] - previous approaches, giving the conditionals a subjunctive interpretation and using known connections between the conditionals and belief revision operators, have been taken by Kern-Isberner and Wobcke [4]. These approaches faced a problem of defining revision operators that avoid Gärdenfors' *impossibility result* [9], that arises due to the *Ramsey Test* ($\phi > \alpha \in K \Leftrightarrow \alpha \in K + \phi$) and the *preservation criterion* (effectively equivalent to Vacuity) being inconsistent with each other for non-trivial cases [10]. Cross and Thomason restrict the revision procedure and show a theory of conditionals that satisfy the Ramsey Test can be found [10]. In contrast, Casini et al. [4] do not give the conditionals a subjunctive interpretation, and use conditional knowledge bases. Having the conditionals as the objects of the belief change implies that [4] does not have the impossibility problem. Other approaches to revision operators include a system of spheres [19].

Belief revision systems are defined by Martins and Shapiro [31] as Artificial Intelligence programs dealing with contradictions. Both theoretical studies and practical implementations have been performed [31]. Research in this area has to address several problems: the inference problem, the non-monotonicity problem, dependency recording, disbelief propagation and the revision of beliefs. Martins and Shapiro [31] elaborate on these and explain the issues involved in each.

Regarding the belief revision postulates or properties, there exists more than one function $*$ that satisfies the AGM properties of belief revision [38]. This is not, however, a weakness - Peppas [38] argues that it simply expresses that people may change their minds in different ways to one another.

2.2.4 Examples. Gärdenfors [10] notes three main methodological questions to resolve: the representation of beliefs in the knowledge base, the relation between explicitly represented elements and derived beliefs, and the decision process regarding what to retract. He illustrates the importance of these with an example [10]: consider a knowledge base that includes information α (All European swans are white), β (The bird caught in the trap is a swan), γ (The bird caught in the trap comes from Sweden) and δ (Sweden is part of Europe). A logical inference from this information is belief ϵ that the bird caught in the trap is white. Suppose we receive the fact that the bird caught in the trap is black. We would want to add $\neg\epsilon$ to the knowledge base, but this would result in an inconsistent collection of information, so a decision is necessitated regarding choosing what propositions to retract prior to adding $\neg\epsilon$. In revision situations, one idea is that information loss from revisions should be minimal whereas another idea is that some beliefs are deemed more entrenched than others and so the least important ones should be retracted [10]. In this example, we can retract α, but then must decide which of its logical consequences we wish to retain.

Another example is given by Casini et al. [4]. In this example, we have an alphabet A = {a, m, n, v}. Respectively, these propositions represent being an avian red-blood cell, being a mammalian red-blood cell, being a vertebrate red-blood cell, and having a nucleus. Consider the situation where knowledge base $K = \{v \vdash n, a \vdash v, m \vdash v, m \vdash \neg n\}$ and $\neg m \in C_n(K)$. The presence of $\neg m$ in $C_n(K)$ when mammalian red-blood cells exist, leads to a conflict. The response is to revise the knowledge base. In the framework that Casini et al. [4] propose, the conflict can resolved by weakening [40] the conflicting proposition(s) already in the knowledge base. In propositional belief change, the conflict would be resolved by eliminating some information, likely either $m \vdash v$, $m \vdash \neg n$ or $v \vdash n$.

Consider a murder trial with α and β our primary suspects, and initial belief base $K = \{(\alpha \wedge \neg\beta) \vee (\neg\alpha \wedge \beta)\}$. We believe one person committed the crime, and we believe it was either α or β i.e. if one of them is innocent, then the other is guilty. During the trial, testimonies are received that incriminate first α and then β. Given that we believe $\alpha \Rightarrow \neg\beta$ and $\beta \Rightarrow \neg\alpha$, the testimonies yield $(K * \alpha) * \beta \models \neg\alpha$ i.e. we believe β committed the murder. In this example, the order in which the information is received affects our final beliefs. This is a problem noted in cases of belief revision iteration [6].

2.3 Belief Update

Belief update is a belief change operation, as is belief revision [13, 18, 19, 24]. As such, in belief update as in belief revision, we assume a classical knowledge base and a classical propositional language, with a Tarskian consequence relation $C_n(\cdot)$. In the event of new information or input corresponding to a change in the world, an update operation is performed [14]. Else, a revision operation is performed [14]. This can be further clarified: the agent's interpretation of the new information is what determines the choice of operation to perform. If the new information is interpreted as an indication the world has changed i.e. there is a dynamic state of affairs, then the choice is belief update. If the new information is interpreted to indicate that the information previously known must be incorrect or flawed i.e. there is a static state of affairs, then the choice is belief revision. New information μ can thus understood as an action effect [24], of which there are two possible types: an *ontic (physical) effect* and an *epistemic effect*. Belief update as a operation can be understood as a form of action progression [24]. The belief change or theory change operation called erasure is to belief update as contraction is to belief revision. In the event of new information μ, the belief update approach would be to take each model of knowledge base or belief base ψ and update it to be a model of μ by as minimal a change as possible [18].

2.3.1 Notation and terminology. We denote an update operation as $\diamond$, defining it as a function accepting input μ, to be applied to a belief base ψ to yield a new belief base $\psi' = \psi \diamond \mu$. The terms *belief base* and *knowledge base* are used interchangeably, as are the terms *belief* and *theory*. Updates are performed world by world [14]. In Section 2.1.1, we defined world and gave notation for models. In belief change literature, $[[\delta]]$ is the preferred notation for the models of sentence δ, where belief bases can be represented by sentences, as can new information. Using this notation, we have $\psi \in [[\delta]]$ to express that ψ is a model of δ i.e. a world in which δ holds. The term *ontic* means feedback-free [24] and *epistemic* means relating to knowledge or to the degree of validation of knowledge.

2.3.2 Properties. The following are the properties of Belief Update as defined by Katsuno and Mendelzon [18] and characterise more than one such operator [14].
1. Success: $\psi \diamond \mu \models \mu$
2. Vacuity: If $\psi \models \mu$ then $\psi \diamond \mu \equiv \psi$
3. Consistency: $\not\models \neg\psi$ and $\not\models \neg\mu$ then $\not\models \neg(\psi \diamond \mu)$
4. Irrelevance of Syntax: $\psi_1 \equiv \psi_2$ and $\mu_1 \equiv \mu_2$ then $\psi_1 \diamond \mu_1 \equiv \psi_2 \diamond \mu_2$
Properties 5-8 do not have titles.
5. $(\psi \diamond \mu) \wedge \phi \models \psi \diamond (\mu \wedge \phi)$
6. If $\psi \diamond \mu_1 \models \mu_2$ and $\psi \diamond \mu_2 \models \mu_1$ then $\psi \diamond \mu_1 \equiv \psi \diamond \mu_2$
7. If ψ is complete, $(\psi \diamond \mu_1) \wedge (\psi \diamond \mu_2) \models \psi \diamond (\mu_1 \vee \mu_2)$
8. $(\psi_1 \vee \psi_2) \diamond \mu \equiv (\psi_1 \diamond \mu) \vee (\psi_2 \diamond \mu)$

2.3.3 Discussion. The difference between the belief update and belief revision operations is first noted by Katsuno and Mendelzon [18], in the context of extended relational databases [14]. Keller and Winslett termed the belief update operation as *change-recording* and the belief revision operation as *knowledge-adding* [18]. Katsuno and Mendelzon [18] extend Keller and Winslett's work [21], formalising their informal distinction of the two operations, working from a more generalised setting and considering a more extensive set of cases. Katsuno and Mendelzon then take it further, propounding a way to combine belief update and belief revision into one operator parameterised by time [18]. Lang [24] uses this idea of time as a parameter in his proposal for belief update as an action progression and new information as an action effect. Boutilier combines belief update and belief revision too, based on a propositional framework [14]. Other approaches to belief update include frameworks using situation calculus and theories of action as the basis [7, 14, 24].

Papers such as [14] argue that properties or postulates 2, 5, 6, and 7 are controversial and should not necessarily be required of belief update operators, and further explore an additional property. Herzig and Rifi [14] break down the KM properties into sub-postulates to identify from where the controversy stems. They also examine ten update operations from the literature, characterising them in terms of strength and computational complexity for comparison purposes and evaluating them with regard to the KM properties. In a critique of [18], they find that two of the ten update operations in literature satisfy their criteria. Herzig and Rifi [14] also study the role and issues of integrity constraints in belief change.

2.3.4 Examples. [18] offers us the following example: the agent has an initial knowledge base ψ comprising that there is either a book (b) or a magazine (m) on the table, but not both. New information is later received - there is a book on the table (b). The revision approach would be for the agent to interpret the information simply as more information about a fixed state of events, and thus for the agent to conclude that there is thus no magazine on the table ($\neg m$). The update approach would be for the agent to interpret the information as that, since receiving the initial information, a book has been placed on the table, The agent thus cannot conclude that, given there is a book on the table, there is no magazine on the table ($\not\models \neg m$). This example highlights that when performing revision, it is because new information about the world has been received, whereas when performing update, it is because the world has changed, casting prior knowledge of the world into uncertainty.

2.4 Defeasible Reasoning

Defeasible reasoning is a form of non-monotonic reasoning. Ragni [40] differentiates strict and defeasible knowledge as the latter allowing for exceptions where the former does not, be it an explicit or implicit exception. Regarding defeasible reasoning, this is possible because the preconditions in a defeasible knowledge base are assumed to hold in the absence of explicit contradictory knowledge [27], which is to say that they are not explicitly true.

2.4.1 Notation and terminology. In defeasible logic, the monotonicity principle of classical logic is weakened [40] and used as one of the properties which all defeasible operations must satisfy. There are eight main properties, and these are formalised and propounded by Kraus, Lehmann and Magidor. [23]. The weakening of the principle of monotonicity is reflected in the notation used for defeasible logic. The concepts of *entailment* and *consequence* adjust to accommodate conclusions being retractable: $\models$ becomes $\approx$ and $\vdash$ becomes $\mid\sim$, where $\alpha \mid\sim \beta$ reads α "typically $\Rightarrow$ " β.

2.4.2 Properties. The following are the properties of Defeasible Reasoning as defined by Kraus, Lehmann and Magidor. [23].
1. Reflexivity (Ref): K $\approx \alpha \mid\sim \alpha$
2. Right weakening (RW): K $\approx \alpha \mid\sim \beta$ and $\alpha \models \gamma$ $\therefore$ K $\approx \alpha \mid\sim \gamma$
3. Left logical equivalents (LLE): K $\approx \alpha \mid\sim \gamma$, $\beta \models \alpha$ and $\alpha \models \beta$ $\therefore$ K $\approx \beta \mid\sim \gamma$
4. And: K $\approx \alpha \mid\sim \beta$ and K $\approx \alpha \mid\sim \gamma$ $\therefore$ K $\approx \alpha \mid\sim \beta \wedge \gamma$
5. Or: K $\approx \alpha \mid\sim \gamma$ and K $\approx \beta \mid\sim \gamma$ $\therefore$ K $\approx \alpha \vee \beta \mid\sim \gamma$
6. Monotonicity: K $\approx \alpha \mid\sim \beta$ $\therefore$ K $\approx \alpha \wedge \gamma \mid\sim \beta$
Classical propositional logic is monotonic. Lehmann and Magidor [26] argue that defeasible entailment ($\approx$) ought to require weaker forms of monotonicity, namely Cautious and Rational Monotonicity.
Cautious Monotonicity (CM): K $\approx \alpha \mid\sim \beta$ and K $\approx \alpha \mid\sim \gamma$ $\therefore$ K $\approx \alpha \wedge \gamma \mid\sim \beta$
Rational Monotonicity (RM): K $\approx \alpha \mid\sim \beta$ and K $\not\approx \alpha \mid\sim \neg\gamma$ $\therefore$ K $\approx \alpha \wedge \gamma \mid\sim \beta$

2.4.3 Discussion. Within defeasible reasoning, there are two different stances on drawing conclusions in the case of exceptions: prototypical reasoning and presumptive reasoning [25]. In prototypical reasoning, given a typical situation, inheriting properties is fine. This type of reasoning is formalised by Lehmann and Magidor [25] using the rational closure defined in an earlier paper of theirs [26]. In presumptive reasoning, if there is no evidence indicating otherwise i.e. if not explicitly negated, inheriting properties is fine. This type of reasoning is the one intended by default logic and is formalised by Lehmann [25]. Logic formalisations such as those used by Lehmann [25] and Magidor [26], focus on the form of the propositions over their meaning, in acknowledgement that the reader's knowledge of the world may influence the study of the formal properties [25]. Pelletier and Elio [37], in contrast, focus on the meanings over the forms. The latter approach does enhance readability, but loses the concise precision of the former and introduces potential variation in the impact on the reader.

2.4.4 Examples. An illustration of the need for defeasible reasoning is given by Kraus, Lehmann, and Magidor [23], by means of the following example: (i) Birds fly ($b \Rightarrow f$), (ii) penguins are birds ($p \Rightarrow b$) and (iii) penguins do not fly ($p \Rightarrow \neg f$). Suppose we have a knowledge base K=(i),(ii). Receiving the additional information that penguins do not fly introduces an inconsistency, or rather an incoherence [4], in that if birds fly and penguins are birds but do not fly, penguins cannot exist ($\neg p$). The solution proposed by [23] is to weaken the proposition that *birds fly* ($b \Rightarrow f$) to *typically, birds fly* ($b \mid\sim f$). There are properties in addition to those given in Section 3.2, which are also desirable [25]. Four such properties that pertain to closure are of particular interest: the presumption of typicality, the presumption of independence, priority to typicality, and respect for specificity. Explanations and justifications of the above properties are given by Lehmann [25].

Reasons need not be beliefs. For example, perceptual states can be reasons [39]. Pollock [39] illustrates this point with an example from the perspective of the agent: sentence β that *X looks red to me* offers a reason for the belief ϕ that X is red. This is, however, a defeasible reason - for example, consider new information α delivered to the agent by a trustworthy person, that *X is not really red, but it appears so due to the lighting conditions*. The agent's reason to believe that X is red (ϕ) no longer justifies it to hold - α is consistent with β, but $\alpha \wedge \beta$ does not offer a reason to believe ϕ. In this example, β is what is called a *prima facie* reason and α is called a *defeater*. Pollock [39] defines and explores different types of defeaters, arguing that prima facie reasons and defeaters are the primary cause of the non-monotonic nature of human reasoning.

2.5 Overview

Belief revision allows for the consideration that to the average human, one belief may have a different level of importance than another belief [38]. The process of the operation reflects that, in the face of new information, the agent may need to change their beliefs and accept the implications the new information may have for their beliefs.

Classical reasoning is deductive i.e. monotonic, whereas non-classical reasoning showcases a flexibility, characteristic of cognitive reasoning, which can be classified as non-monotonicity. For each of the non-classical forms of reasoning we considered, we assumed an underlying propositional logic. Conflicting information is interpreted, and thus resolved to preserve consistency, differently in each. In defeasible reasoning, conflicting information generally indicates exceptions - facilitated by means of weakening [40] notation, reflecting a more relaxed monotonicity principle than that prevalent in classical reasoning. This manifests through the notion of *typicality*. Both belief revision and belief update are considered to be forms of belief change. Belief change involves a belief base and a belief set, where explicit knowledge resides in the base and inferences or knowledge derived from that in the base resides in the belief set. In belief revision, conflicting information indicates flawed prior knowledge on the part of the agent, defeating, i.e. forcing the retraction of, conclusions drawn from it. Such information is referred to as a defeater, either of type *rebutting* or *undercutting* [39], and is taken into account by selecting the models of the new information closest to the models of the base [18]. In belief update, the notion of time enters the scenario, contrasting the static environment assumed in belief revision to a dynamic environment [18] assumed in belief update. This means that conflicting information is taken to indicate or correspond to a change in the world or real state of affairs.

3 DESIGN AND EXECUTION

3.1 Experiment design, execution and challenges

The experiment involves developing questions, creating a survey using Google Forms and requesting responses via a platform offered by Amazon, called Mechanical Turk (MTurk).

3.1.1 Questions. The questions were developed to test whether properties of a specific formalisation of the process of Belief Revision feature in cognitive reasoning. The formalisation used is that of the eight-postulate approach as proposed by Alchourrón, Gärdenfors and Makinson (AGM) [1]. Two types of questions were developed: concrete and abstract. This involved designing scenarios in which to ground the concrete questions. Five such scenarios were designed, with ten concrete questions being formulated overall. Eight abstract questions were developed, directly based on the formal properties. The abstract questions were included so as to test the properties without having the agent's knowledge of the world hindering their answers and to have questions which are less semantically loaded [25] than real-world concrete questions. The benefit of abstract examples is further discussed by Pelletier and Elio [37]. The concrete questions started out as abstract representations explicitly requiring the application of one or some of the formal properties to obtain the desired answer. These representations were then elaborated in the context of a scenario. An example of a concrete, story-style or real-world question would be: "If Cathy has a cake to bake, will she use an oven?". An example of an abstract question would be: given the following, "If A then B", and "If C then A", can we say that "If C then B?".

The scenarios designed are: linguists, smoking, wildlife, bag of stationery and, acrobats. The scenarios designed are partly inspired by the literature discussed in Section 2 and partly by the researcher's knowledge of the world and creativity. The linguist scenario is inspired by the former and the other four scenarios are inspired by the latter. The wildlife scenario thus reflects a South African influence, through the inclusion of a giraffe.

3.1.2 Survey. The survey was constructed using Google Forms and comprises eight sections. The first section offers the participant a short explanation regarding the purpose of the survey. Given that the survey is answered online and not in person, the second section is a checkpoint, designed to be an indicator of the suitability of the respondent to take the survey. In this context, suitability comprises four requirements: *(i)* the response is not generated by a bot, *(ii)* the respondent is not using a script, *(iii)* the respondent can understand English, *(iv)* the respondent reads questions in full.

Research performed in 2018 revealed that the recent surge of low-quality qualitative data from MTurk is primarily due to international Turkers (workers on MTurk) [43] using Virtual Private Networks (VPNs) or Virtual Private Servers (VPSs) to waive qualifications required to complete surveys [22]. This motivated including a checkpoint within the survey itself. The checkpoint comprises two questions. To test suitability, the first question is designed in the style of a captcha, with the purpose of a honeypot [43, 44]. The respondent is presented with three images, each of a shape. The official instruction is to select the image showing an open figure, but in the question blurb, the respondent is instructed to ignore the question and not give an answer, where failure to comply means disqualification from the survey. The second question is discussed in Section 3.1.3.

The third section informs the participant about the research study, their role and rights in the process, provides the contact details of the researcher and the Computer Science department, and asks the participant to indicate if they feel sufficiently informed and voluntarily consent to taking part in the survey. Participants that do give their consent are directed to the next section; participants that do not are directed back to the first section of the survey. The fourth section presents a brief background of Belief Revision, laying the foundation for the instructions given in the next section. The fifth section contains the questions - 18 in total. At the start of the section, there are instructions regarding how to read the questions and what not to do when answering the questions. Specifically, the participant is told that there are no right or wrong answers, so they should not hold back when giving their explanations. The sixth section asks the participant about themselves, i.e. what their relationship with the survey is and into what age category they fall, and also asks for feedback on the survey. This is an important section, as it informs the analysis of the data gathered, giving insight into the participant and their experience of the survey process.

The last two sections in the survey work together regarding tracking responses. The survey was posted on Amazon's MTurk as a *Human Intelligence Task* (HIT) using a link to the Google Form. For HITs defined in this way, MTurk requires respondents to enter a completion code in the HIT to verify that they have successfully completed the survey. Matching a response to a *Worker*, also called a *Turker*, on MTurk can be done either by using a unique completion code for each response or by requesting the Worker's ID in the survey [33]. The completion code in this case is not unique to each response; it is held constant. The seventh, or penultimate, section thus asks the participant to enter their Worker ID and the final section specifies the only correct completion code to be used.

3.1.3 Google Forms. Google Forms is a survey administration application that allows users to create free online surveys. To prevent the survey from being spammed with responses, a captcha was included as part of the checkpoint section discussed in Section 3.1.2. Google Forms does not currently support recaptcha, so whilst the captcha is outdated as a means of bot detection, one offered by *xFanatical* [15, 16] is included as a precautionary measure in the survey.

3.1.4 Mechanical Turk. Mechanical Turk (MTurk) is a service provided by Amazon that serves as an interface between *service requesters* and a network of humans. It addresses three problems [17]. It is used by software developers to incorporate human intelligence into software applications. It is used by business people to access a large network of human intelligence to complete tasks, for example to conduct market research. It is used by people looking to earn money to find work that can be done anywhere and at any time, using the skills they already have. In this regard, the idea was to create a space for people to earn money in their spare time [17], but it has since evolved into a platform used by many as their primary source of income [42]. This introduces ethical issues regarding minimum compensation amounts.

This project used MTurk for access to its network of humans to complete a survey hosted on Google Forms. The advantage of MTurk is that its network of Workers includes people from many places, with a large range of ages and education levels [41]. Such places include the United States of America, Canada, India, Pakistan, the United Kingdom and the Philippines [41]. The reward per response should have a lower bound of the minimum wage, for ethical reasons [42]. The location of respondents requested for this project, however, was not restricted. This meant that the minimum wage in South Africa (R20/hour [34]) could not be used as the lower bound, as there are other countries with greater amounts as their minimum. With project feasibility in mind, in terms of funding, a compromise was made: to ensure the compensation offered was greater than the South African minimum wage, and rely on the comparatively low amount to deter those for whom MTurk is their primary source of income from accepting the task.

Another ethical consideration with MTurk is the data involved. HITs may not contain personal or sensitive data and service requesters may not publish HITs that can derive personally identifiable information from the respondent [17]. The questions in the survey for this project do not seek such information; their focus is how the respondent reasons. In the feedback section of the survey, respondents are asked to select the age bracket in which they fall and their Worker ID. The former is used as is to inform the analysis, whereas the latter is used to match a Worker requesting compensation to the corresponding HIT response. Neither is used to derive further information that could reduce the anonymity or invade the privacy of the respondent.

Publishing the final survey entailed posting the survey link as part of a HIT on the site. The HIT was created with certain specifications accordingly. Six batches of the HIT were published, with each batch of the HIT requesting between 4 and 7 responses. The varied number of responses requested was due to batch size restrictions given rejections of unsuitable responses. The HIT is titled *Answer a survey about how you reason* and described as requiring responses to given scenarios in the face of new information. To assist Turkers finding the HIT, it has the assigned keywords *survey, human, reasoning, academic study*. The reward per response is set to $2.70. The expected average time to complete the survey was 30-40 minutes, based on the average time taken to complete the trial survey. For a discussion on the trial survey, refer to Section 5.1. Workers are thus given an hour to complete the survey, to avoid having rushing influence answers. The survey's lifetime was set to four days, with auto-approval set to occur three days after reception of a response. Three restrictions, in the form of qualifications, were enforced regarding workers that may take part in the survey. Two qualifications are system qualifications: Workers were required to have a HIT Approval Rate (%) for all Requesters' HITs of greater than 98, and have more than 5000 HITs approved. The required number of HITs approved was varied, as 1000 and 5000, to allow for a more diverse sample of respondents. It was reasoned that perhaps Turkers with fewer HITs approved may also be those that have answered fewer HITs due to being in a different life situation, for example having a full-time job outside of being a worker on MTurk. One qualification was created by the researcher, to ensure that the 30 respondents were unique across all of the published batches of the survey. This qualification is called *Completed my survey already* and is assigned to Workers which have submitted a response in a previous batch, including the batch of the trial HIT.

3.2 Testing and evaluation

The survey, once constructed, was tested. This entailed evaluation of the survey by a group of laypeople and experts and the publication of the survey in a HIT on MTurk as a trial. The results of the trial HIT were evaluated, to gauge both how Turkers might respond to the final survey and the results that may be expected, the latter in terms of belief revision properties being employed in human reasoning.

3.2.1 Feedback from group of laypeople and experts. The survey was evaluated by 4 people: Mr Clayton Baker and Mr Paul Freund, as informed laypeople; Professor Thomas Meyer, as an expert within the Computer Science Department; a Computer Science Honours student with no involvement in this project, who wishes to remain anonymous, as a layperson.

Mr Clayton Baker and Mr Paul Freund were conducting similar studies to this one at the time, focusing on Defeasible Reasoning and Belief Update respectively. They were sent a Google Forms link to the survey, with the request to answer it providing feedback on both the questions and the overall survey experience. They identified that the second question of the checkpoint in the survey was refreshing too frequently, making it difficult to pass the test. This was adjusted accordingly. Instructions were found to be clear and coherent and questions were deemed to be fine overall. A comment was made that the wildlife scenario caused their intuition to blur. This was not unexpected, as the scenario presents information not aligned with reality. A suggestion was made to remove from the consent form the mention of reasoning in AI systems other than Belief Revision. Another suggestion was to include a progress bar, ensuring respondents are kept informed regarding how much of the survey they have still to complete. Both suggestions were taken. The abstract questions were found to be more challenging to understand than the concrete questions in terms of what is being asked, particularly in cases where the question asks the reasoner to consider "the reverse" scenario of another question.

Feedback from Professor Thomas Meyer and Mr Zola Mahlaza was sought after updating the survey based on the feedback from the informed laypeople. The questions were deemed to be fine, although several comments were made regarding how they could improve. The questions involving the keyword "*or*" could be found to be ambiguous, so the desired interpretation should be specified for clarity. Furthermore, the punctuation involved in presenting the scenarios and new information should have their use explained. The abstract questions referring to "*the reverse*" of another abstract question should include a line of elaboration for clarification. The survey was updated accordingly.

Throughout the design phase of the project, informal evaluation was requested from other Computer Science Honours students, selected by convenience. This was done informally, but provided invaluable clarity-and-coherence checking of the instructions and of the survey overall.

3.2.2 Trial HIT. A trial of the survey using MTurk was conducted. This was done *(i)* to gain familiarity with the MTurk service and platform and *(ii)* to test the survey and its questions on a sample of Turkers.

It involved posting the survey link as part of a HIT on the site. The HIT was created with certain specifications accordingly. The trial survey is titled *Answer a survey about how you reason* and described as requiring responses to given scenarios in the face of new information. To assist Turkers finding the HIT, it has the assigned keywords *survey, human, reasoning, academic study*. The reward per response is set to $2.12, with five respondents requested. The expected average time to complete the survey was 20 minutes. Workers are thus given 40 minutes to complete the survey, to avoid rushing influencing answers. The survey's lifetime was set to four days, with auto-approval set to occur three days after reception of a response. Two restrictions, in the form of qualifications, were enforced regarding workers that may take part in the survey. Workers were required to have a HIT Approval Rate (%) for all Requesters' HITs of greater than 98, and have more than 5000 HITs approved.

Six responses were received, five of which are discussed. One of the first five responses failed the checkpoint, so is not included in the evaluation. Of the five remaining workers, one did also fail the checkpoint but, given the quality of their written answers, this is deemed likely a result of not reading the instructions thoroughly only for the beginning of the survey. All workers gave their consent to take part in the survey, and on average took 28 minutes 23 seconds to complete it. This was greater than the expected average time of 20 minutes. Compensation or reward per response was increased to $2.70 to reflect both this and the open questions added for the abstract questions in the final survey. Allotted time to complete the final survey was increased to an hour, with the expected average time adjusted to 35 minutes.

4 METHODS OF ANALYSIS

Analysis of the Questions section of the survey, for both the trial and final survey, comprised finding the modal answer and hit rate for each closed question and performing qualitative analysis on the open questions.

The data was downloaded from Google Forms and Mechanical Turk. The modal answer and hit rate (%) for closed questions were determined by applying functions in Microsoft Excel to the data. A hit indicates success. In this context, success is defined as both the respondent and the application of the properties of belief revision obtaining the same answer for a question. Hit rate is thus calculated for each question as $\frac{number of successes}{no. of responses} \times 100$.

The qualitative analysis was performed in NVivo, a qualitative data analysis software package, and made use of *Tesch's Eight Steps in the Coding Process* [5]. In this context, coding refers to a process of organising data, specifically by bracketing segments of data into themes, or topics. These are then given labels i.e. codes. Themes are, in turn, bracketed into categories, which are then labelled.

In this process, a combination of pre-determined and emerging codes were used. Codes on topics expected to be found were taken from literature, based on the theory being empirically tested. These include the eight properties of belief revision as proposed by Alchourrón, Gärdenfors and Makinson [1]: closure, success, inclusion, vacuity, consistency, extensionality, super-expansion, sub-expansion. Other pre-determined codes include: normative and descriptive. Emerging codes are those which were not anticipated at the beginning, or are both unusual and of interest. They are developed solely on the basis of the data collected from respondents by means of the survey. An example of an emerging code used in the trial of this study is *It is stated*. This code represents the respondent taking a passive approach to their response. Other examples would be *real-world influence* and *likelihood*.

Pre-determined codes *normative* and *descriptive* refer to the reasoning style identified in responses to open questions. A normative style involves making value judgements [35], commenting on whether something is the way it should be or not. This includes implied judgements through the use of emotive language. A descriptive style, in contrast, does not - it involves making an observation, commenting on how something is [35].

5 RESULTS AND DISCUSSION

A high-level analysis of the trial survey was performed to inform our expectations for the final survey.

5.1 Trial survey

Four of the five workers said they took part because of general interest, and three of the five workers said they took part to earn money. Three of the workers were between ages 20-35, and the other two were older, between ages 45-55. All five workers said they found the instructions to be clear and coherent and the background information helpful.

It can be seen in Figure 2, that overall the hit rates are higher for the real-life concrete questions (Q1-Q10) than the abstract questions (Q11-18). This supports the idea that the grounding of questions in scenarios more familiar and story-like makes them clearer for the participant to reason within and understand. One exception would be Question 7, which has a hit rate of 0%. It is worth noting that the scenario in which Question 7 is set directly conflicts with real-world knowledge. Explanations given for answers to Question 7 indicate confusion and surprise with regard to the information received.

The answers to the open questions revealed themes which fall into four main categories : new information, the survey, known influences and capacity. Regarding new information, it was identified that in the majority of questions, the belief revision property of success is employed. New information was found to be rejected when believing it is deemed "naïve", indicating there may be a degree of epistemic entrenchment at play. New information was also rejected when deemed "confusing" and illogical. Regarding the survey, it was identified that the inconsistent structure of the questions, i.e. some questions specifying two different cases, caused confusion as some responses received assumed dependence between the two. Some syntax used, such as the use of "OR" in questions testing *Inclusion, Vacuity, Sub-expansion* and *Super-expansion* or the use of a double negative used in a question testing *Extensionality*, was not understood. In 3 out of 5 responses, the double negative was interpreted as a single negative. Regarding known influences, it was found that both calculated and perceived likelihood played a role in people's responses, as did real-world influence. Real-world influence was noted in several forms, such as one respondent saying that reasoning in the manner required was unfamiliar, as usually they search the internet for answers to things of which they are uncertain. It was also noted in that people rely on their real-world knowledge to answer questions, sometimes bringing in their own new information to answer the question, which let them down in Question 7, as responses included comments saying they "clearly do not know enough about giraffes". Regarding capacity, it was identified that some responses increase in brevity with an increase in confusion or clarity on the part of the reasoner. This would suggest that their capacity for patience and willingness to explain answers thoughtfully decreased in these cases. Another observation is that there was a marked lack of closure, i.e. thinking through of consequences, evident in the responses. Possible reasons for this could include: workers on MTurk not typically being asked to answer questions that require them to reason in such a manner, workers on MTurk having less general interest in the survey than their interest in earning money, people not wanting to do extra work revising their beliefs until necessary i.e. taking a lazy-evaluation style approach to their beliefs and, people not realising the consequences of the new information received.

Of the five respondents, four were classified as having a descriptive reasoning style and one as having a normative reasoning style.

5.2 Final survey

Forty Turkers responded to the HIT, ten of which failed the checkpoint section of the survey and are thus excluded from the analysis. The Google Form received 42 responses, as two Turkers submitted twice. The duplicates were discarded. All workers gave their consent to take part in the survey, and on average took 42 minutes to complete it, exceeding our expectation by seven minutes.

5.2.1 Expectations. MTurk is a formal platform enabling workers to complete tasks for compensation. We thus expected earning money to be the primary objective of the MTurk workers i.e. *Turkers* that respond to our survey. MTurk does not accept registrations from workers less than eighteen years of age. There was no expectation regarding age brackets of the respondents in this study beyond that they would not be minors. Based on the trial survey, we expected respondents to find the instructions clear and coherent, and the background information clear and helpful. Based on the trial survey, we expected the concrete questions to have, overall, higher hit rates than the abstract questions. Confusion was anticipated regarding Question 7. Question 3 and 7 were expected to have the lowest hit rates, as in the trial survey. We expected descriptive reasoning to be more prevalent than normative reasoning in respondents, based on both our trial survey and Besold and Uckelman's arguments [3].

5.2.2 Qualitative and quantitative analysis. 27 of the respondents said they took part to earn money, 15 because of general interest and three said they are researchers. This matches our expectation of earning money being the primary objective. For Turkers, time spent taking longer on HITs typically means achieving a lesser monetary throughput. In the context of our survey, this does introduce some bias into our sample of respondents, where respondents may be less inclined to give the questions thorough thought and share their insights into their reasoning process. We make the assumption that Turkers value their qualifications, for example measured by their *HIT approval rate*, considering this to partially offset strong bias regarding compensation.

The modal age category was 25-30 years, with the majority (28) of the respondents being less than 40 years of age. This, paired with the feedback received regarding the relevance of the survey to respondents, would suggest that the respondents are not full-time Turkers. Respondents' fields of work were varied and include computer science, retail, teaching, management and law.

21 of the 30 respondents (70%) found the instructions to be clear and coherent, and 24 (80%) found the background information helpful. Most of the respondents (26 out of 30) said that they reasoned similarly to their day-to-day life when answering the concrete questions. In contrast, 19 of the respondents found the abstract questions to be understandable. This supports the idea that the grounding of questions in scenarios more familiar and story-like makes them clearer for the participant to reason within and understand. It is reflected in Figure 1, which illustrates that the hit rates for Success, Closure, Sub-expansion, Super-expansion and Consistency are greater for the concrete questions than the abstract questions. The hit rates for Extensionality, Inclusion and Vacuity are greater for the abstract questions than the concrete questions.

A closer investigation Figure 1 reveals that Question 7 has the lowest hit rate (13.33%). At a first glance, this is consistent with the trial survey findings, but the causes differ. Story-style or concrete questions, whilst more readable, can be confusing when the scenario does not align with reality. This was certainly the case in the trial survey, and it manifest in a hit rate of 0%. Whilst the scenario in question did not change from the trial to the final survey, respondents to the final survey largely accepted the contrary information regarding giraffes and rather interpreted the terminology used ("OR") in different ways. This was not unexpected, as semantically *or* has two interpretations (*inclusive* or *exclusive*), whereas disjunction has one. Those that did not accept the misalignment with reality brought in their own information to answer the question, most turning to results from a related Google search. This was a common occurrence throughout the open questions in the survey. In a world where instant gratification regarding knowledge has become an almost ubiquitous option given online search engines such as Google, this finding may suggest that the way in which the typical human reasons is changing. For example, it could suggest that in the face of conflicting information, cognitive reasoning is becoming less dependent on the human agent, or reasoner, and more dependent on the technology that answers their questions, effectively outsourcing the work. The postulate *Closure* (see Section 2.2.2) refers to logical omniscience. In the case of relying on technology for answers, such a vast amount of information regarding the real world at one's fingertips may bring humans closer in their daily decision-making, albeit artificially so, to the ideal that closure presents. Question 3 has the second lowest hit rate, as can be seen in Figure 6, at 26.67%. The corresponding open question revealed that the use of a double negative in Question 3 was not understood. In 10 out of 30 responses, it was clear that the double negative was interpreted as a single negative. An additional 10 respondents did not register that there was a double negative in play. Such results may well be due to different dialects, different education levels, or, simply the level or style of English among respondents. Another cause could be respondents' perceived level or style of English of the researcher.

The answers to the open questions revealed themes which have been categorised as follows: *Pre-determined* and *Emergent*. Pre-determined comprises *Reasoning Style* and *Postulates*. Themes that emerged during analysis are: rejection of new information, known influences, expression affects interpretation and, capacity. New information was found to be rejected because a new, more specific belief was less entrenched than an older, more general belief with which it conflicted. Respondents distrusted new information when they deemed it logically implausible and, in some cases, because they could not evaluate the source of the information. In the event of the question causing confusion, some respondents disregarded the new information altogether. Others claimed to be exercising caution and expressed a preference for the pre-existing consistency of what is known, as they deem it better for inference. It may be that reducing the importance of the new information is their response to cognitive dissonance [32].

Known influences include probabilistic-style reasoning, manifesting in likelihoods being given as motivations for answers. It also includes real-world influence. The concept of time was inserted into seven respondents' reasoning, linking to belief update (Section 2.3). It was noted that four respondents took conflicts as indications of exceptions, linking to defeasible reasoning (Section 2.4). Regarding bringing in own information, it was identified that turning to Google for answers was a popular choice among respondents, as was indulging in confirmation bias.

In questions testing extensionality, responses received commented that success can be largely dependent on specifics of the conflict and the wording. Respondents observed that extensionality is constrained - commenting that whilst syntax can make expression and understanding easier, not everybody can necessarily understand the underlying ideas behind such syntax. Syntax can have different effects on a reasoner's understanding. Case in point, consider the marked lesser comprehension of the abstract questions than of the concrete questions.

Regarding capacity, it was found that question order was muddled by some respondents in that their written explanations sometimes corresponded to a previous answer. Syntax was misunderstood, there was confusion regarding Yes and No, the cases were conflated at

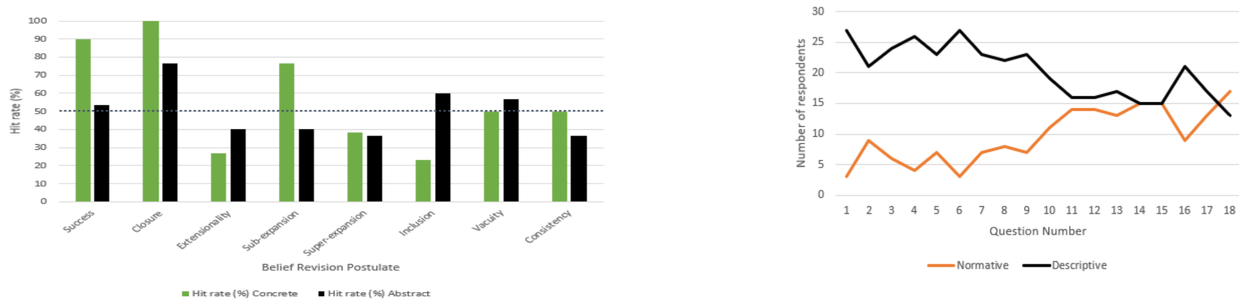

Figure 1: The figure on the left compares the hit rates of the concrete and abstract questions, per postulate. The figure on the right shows the distribution of normative and descriptive reasoning styles, per question. Questions 1-10 are concrete and questions 11-18 abstract.

times, or simply not seen as independent. Other themes in this category include the order of information being deemed irrelevant and respondents valuing being open-minded.

Of the 30 respondents, four were classified as having a normative reasoning style and 24 as having a descriptive reasoning style. The remaining two respondents had an inconclusive reasoning style; they gave the same number of normative answers as descriptive answers. This has been classified as *Other* in the results.

It can be seen in Figure 1 that the concrete questions yielded predominantly descriptive reasoning approaches. This evens out for the abstract questions, with the abstract questions testing Vacuity and Consistency actually both yielding an equal number of descriptive and normative reasoning approaches. The abstract question that caused the most confusion (hit rate 36.67%; corresponding open question taking complaints) was testing super-expansion, and yielded more normative reasoning approaches than descriptive. This could be humans resorting to rules in the face of confusion much like some humans crave routine in chaos.

6 CONCLUSIONS AND FUTURE WORK

The properties of Belief Revision [1] are employed to varying degrees in human reasoning. Success, Closure and Vacuity all have hit rates greater than or equal to 50% for both the concrete and abstract questions. Success has a 90% hit rate in the concrete and 53% in the abstract. It is directly tested in Question 1 and indirectly tested in Questions 2, 3, 8, 9 and 12. Closure was perceived by respondents to be something that they employ in their reasoning. Whether this is true in reality may be a separate issue. Extensionality, whilst better understood in the abstract, did not have a high hit rate. Unlike formalised notation, spoken language does not have a single meaning as the accepted standard for a given expression. The double negative used in the concrete question testing extensionality was misinterpreted by a third of the participants. Sub-expansion and super-expansion manifest more in the concrete than the abstract, as did Consistency. In contrast, Inclusion manifest more in the abstract than the concrete.

In both the trial survey and the final survey, the reasoning style that is the most prevalent is *descriptive*.

We learned that inconsistency in style across questions, beyond the concrete and abstract difference, creates confusion. In future, we would keep the style constant. Given the variation in response quality from Turkers, we would add qualifications to the HIT specification to restrict location or language, or, screen respondents with a pre-survey HIT to ensure they have a suitable level of English. Constructing the survey highlighted the importance of understanding, and defining, all syntax and notation used. Responses taught us the importance of testing. In future, we would test our survey on more than five people. We learned there is a strong reliance on search engines when people have access to online resources. We therefore would use blocks in a future study: *offline* and *online* survey responses. We saw that responses to cognitive dissonance differ, which indicated a link to Peppas' argument [38] supporting the concept of the AGM model [1] characterising more than one such Belief Revision operator.

Two similar projects were undertaken, but with a focus on Belief Update (Section 2.3) [8] and Defeasible Reasoning (Section 2.4) [2] respectively. An avenue that could be explored is a comparison of the findings from these papers. Given that there are close formal parallels between the postulates for belief revision and defeasible reasoning such that that they may be seen as "basically the same process, albeit used for two different purposes" [11], it may be of particular interest to conduct an investigation in that direction. It would be a first step towards testing whether the translations between defeasible reasoning and belief revision hold empirically. For example, if all postulates for defeasible reasoning are empirically found to be employed in human reasoning, but some of the postulates for belief revision are not, given that these postulates are held to be equivalent when viewed in a certain way [12], the results of the projects would suggest that the formal link does not hold empirically.

ACKNOWLEDGMENTS

We would like to Professor Thomas Meyer for his helpful feedback and guidance with the project topic material, and to Mr Zola Mahlaza, particularly for his guidance regarding the project proposal. Thanks are also due to all who participated in the survey, and those who played a role in its evaluation phase.

REFERENCES

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50 (1985), 510–530. https://doi.org/10.2307/2274239

[2] Clayton Baker. 2019. *On the relation between defeasible and human reasoning*. Honours Paper. University of Cape Town.

[3] T. R. Besold and S. L. Uckelman. 2018. Normative and descriptive rationality: from nature to artifice and back. *Journal of Experimental & Theoretical Artificial Intelligence* 30, 2 (2018), 331–344. https://doi.org/10.1080/0952813X.2018.1430860 arXiv:https://doi.org/10.1080/0952813X.2018.1430860

[4] Giovanni Casini, Eduardo Fermé, Thomas Meyer, and Ivan Varzinczak. 2018. A Semantic Perspective on Belief Change in a Preferential Non-Monotonic Framework. In *Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, M. Thielscher and F. Toni (Eds.). AAAI.

[5] John W. Creswell. 2014. *Research design: qualitative, quantitative, and mixed methods approaches* (4 ed.). SAGE Publications, Thousand Oaks, California, 245–253.

[6] Adnan Darwiche and Judea Pearl. 1997. On the Logic of Iterated Belief Revision. *Artificial Intelligence* 89 (01 1997), 1–29. https://doi.org/10.1016/S0004-3702(96)00038-0

[7] Alvaro del Val and Yoav Shoham. 1992. Deriving Properties of Belief Update from Theories of Action. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI'92)*. AAAI Press, San Jose, California, 584–589. http://dl.acm.org/citation.cfm?id=1867135.1867225

[8] Paul Freund. 2019. *Evaluating the Katsuno-Mendelzon Postulates for Belief Update against human reasoning*. Honours Paper. University of Cape Town.

[9] Peter Gärdenfors. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press.

[10] Peter Gärdenfors. 1992. *Belief Revision: An Introduction*. Cambridge University Press, 1–26. https://doi.org/10.1017/CBO9780511526664.001

[11] P. Gärdenfors and D. Makinson. 1994. Nonmonotonic inference based on expectations. *Artificial Intelligence* 65, 2 (1994), 197–245.

[12] S.O. Hansson. 1999. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Kluwer Academic Publishers, Berlin.

[13] Sven Ove Hansson. 2017. Logic of Belief Revision. In *The Stanford Encyclopedia of Philosophy* (winter 2017 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[14] Andreas Herzig and Omar Rifi. 1999. Propositional belief base update and minimal change. *Artificial Intelligence* 115, 1 (1999), 107 – 138. https://doi.org/10.1016/S0004-3702(99)00072-7

[15] Jason Huang. 2019. 3 Ways to Protect Google Forms from Spamming. (07 2019). https://xfanatical.com/blog/3-ways-to-protect-google-forms-from-spamming/

[16] Jason Huang. 2019. Google Forms Captcha: Anti-Spamming for Google Forms. (04 2019). https://xfanatical.com/blog/captcha-for-forms/

[17] Amazon Mechanical Turk Inc. 2018. FAQs. (2018). https://www.mturk.com/help

[18] Hirofumi Katsuno and Alberto O. Mendelzon. 1991. On the Difference Between Updating a Knowledge Base and Revising It. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 387–394. http://dl.acm.org/citation.cfm?id=3087158.3087197

[19] Hirofumi Katsuno and Alberto O. Mendelzon. 1991. Propositional Knowledge Base Revision and Minimal Change. *Artif. Intell.* 52, 3 (Dec. 1991), 263–294. https://doi.org/10.1016/0004-3702(91)90069-V

[20] Hirofumi Katsuno and Alberto O. Mendelzon. 2003. On the Difference between Updating a Knowledge Base and Revising it. *Belief Revision* 29 (2003), 183.

[21] Arthur M. Keller and Marianne W. Wilkins. 1985. On the Use of an Extended Relational Model to Handle Changing Incomplete Information. *IEEE Transactions on Software Engineering* SE-11, 7 (July 1985), 620–633. https://doi.org/10.1109/TSE.1985.232506

[22] Ryan Kennedy, Scott Clifford, Tyler Burleigh, Ryan Jewell, and Philip Waggoner. 2018. The Shape of and Solutions to the MTurk Quality Crisis.

[23] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.* 44 (1990), 167–207.

[24] Jérôme Lang. 2007. Belief Update Revisited. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence (IJCAI'07)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2517–2522. http://dl.acm.org/citation.cfm?id=1625275.1625681

[25] Daniel Lehmann. 1995. Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence* 15, 1 (1995), 61–82.

[26] Daniel Lehmann and Menachem Magidor. 1992. What does a conditional knowledge base entail? *Artif. Intell.* 55 (1992), 1–60.

[27] Luca Longo and Pierpaolo Dondio. 2014. Defeasible Reasoning and Argument-Based Systems in Medical Fields: An Informal Overview. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*. IEEE, 376–381. https://doi.org/10.1109/CBMS.2014.126

[28] David Makinson. 1993. Five Faces of Minimality. *Studia Logica* 52 (09 1993), 339–380. https://doi.org/10.1007/BF01057652

[29] David Makinson. 2005. *Bridges from Classical to Nonmonotonic Logic*. King's College Publications, King's College London.

[30] David Makinson and Peter Gärdenfors. 1991. The relations between the logic of theory change and nonmonotonic logic. In *The Logic of Theory Change*, André Fuhrmann and Michael Morreau (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 183–205. https://doi.org/10.1007/BFb0018421

[31] João Martins and Stuart Shapiro. 1988. A Model for Belief Revision. *Artif. Intell.* 35 (01 1988), 25–79. https://doi.org/10.1016/0004-3702(88)90031-8

[32] Saul McLeod. 2018. Cognitive Dissonance. (02 2018). https://www.simplypsychology.org/cognitive-dissonance.html

[33] /@mechanicalturk. 2019. Tutorial: Getting great survey results from MTurk and Qualtrics. (04 2019). https://medium.com/@mechanicalturk/tutorial-getting-great-survey-results-from-mturk-and-qualtrics-f5366f0bd880

[34] Lameez Omarjee. 2019. Everything you need to know about the national minimum wage. (01 2019). https://www.fin24.com/Economy/everything-you-need-to-know-about-the-national-minimum-wage-20190101

[35] David Over. 2004. Rationality and the Normative/Descriptive Distinction. In *Blackwell handbook of judgment and decision making*, Derek J. Koehler and Nigel Harvey (Eds.). Blackwell Publishing Ltd, United States, 3–18.

[36] Francis Pelletier and Renee Elio. 1997. What Should Default Reasoning Be, by Default? *Computational Intelligence* 13 (05 1997), 165–187. https://doi.org/10.1111/0824-7935.00037

[37] Francis Pelletier and Renee Elio. 2005. The Case for Psychologism in Default and Inheritance Reasoning. *Synthese* 146 (08 2005), 7–35. https://doi.org/10.1007/s11229-005-9063-z

[38] Pavlos Peppas. 2008. Belief Revision. In *Handbook of Knowledge Representation*, F. van Harmelen, V. Lifschitz, and B. Porter (Eds.). Elsevier Science. https://doi.org/10.1016/S1574-6526(07)03008-8

[39] John L. Pollock. 1987. Defeasible reasoning. *Cognitive Science* 11, 4 (1987), 481 – 518. https://doi.org/10.1016/S0364-0213(87)80017-4

[40] Marco Ragni, Christian Eichhorn, Tanja Bock, Gabriele Kern-Isberner, and Alice Ping Ping Tse. 2017. Formal Nonmonotonic Theories and Properties of Human Defeasible Reasoning. *Minds and Machines* 27 (02 2017). https://doi.org/10.1007/s11023-016-9414-1

[41] Joel Ross, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 2009. Who are the Turkers? Worker Demographics in Amazon Mechanical Turk. (01 2009).

[42] Alana Semuels. 2018. The Internet Is Enabling a New Kind of Poorly Paid Hell. (01 2018). https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/

[43] TurkPrime. 2018. After the Bot Scare: Understanding What's Been Happening with Data Collection on MTurk and How to Stop it. (09 2018). https://blog.turkprime.com/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-t

[44] Remya Vinayakumar, Alisha Nelson, Diana P.A., Emilda Jojo, Ranjith R, and Sheril Johnson. 2018. Honeypot Captcha against Spam bots. *International Journal for Research in Applied Science and Engineering Technology* 6 (03 2018), 1010–1014. https://doi.org/10.22214/ijraset.2018.3161

SUPPLEMENTARY INFORMATION

Questions 1-10 are concrete questions; questions 11-18 are abstract.

S1 contains figures, including Figures 5 and 7 which together comprise Figure 1 found on page 10.
S2 contains tables. In this section, more information is given regarding the questions and summaries of the results are provided.

S1 : FIGURES

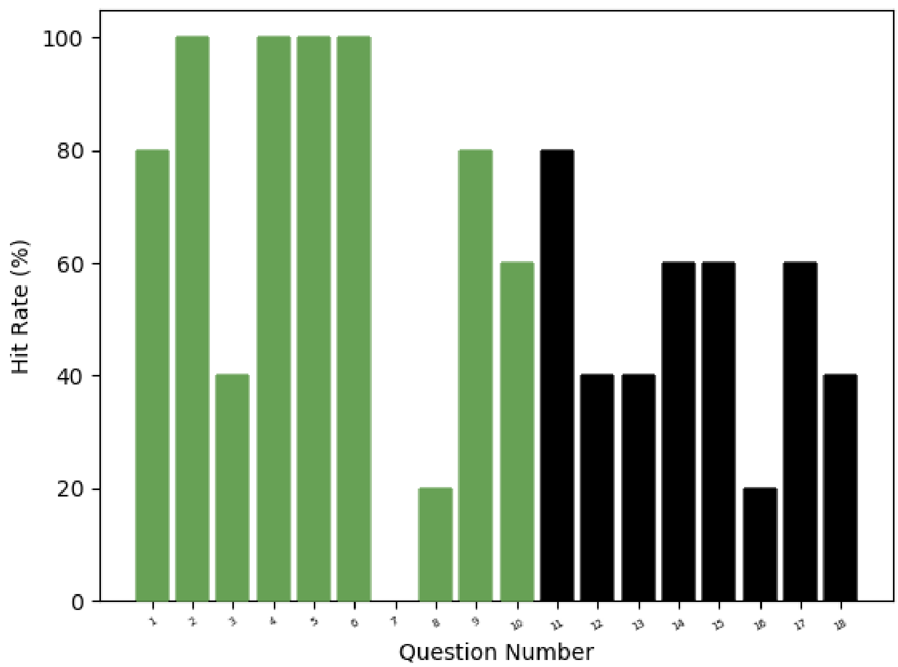

Figure 2: Bar graph showing Hit Rate per question, regarding corresponding answers between Belief Revision and the reasoning of the humans that took part in the trial survey on Mechanical Turk.

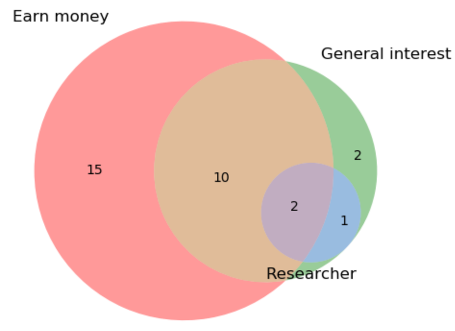

Figure 3: Venn diagram showing the relationship of the humans that took part in the final survey on Mechanical Turk to the survey.

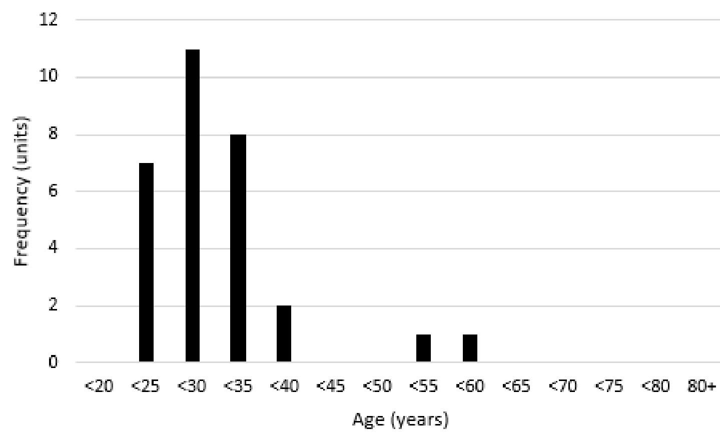

Figure 4: Bar graph showing the distribution of ages of the respondents for the final survey

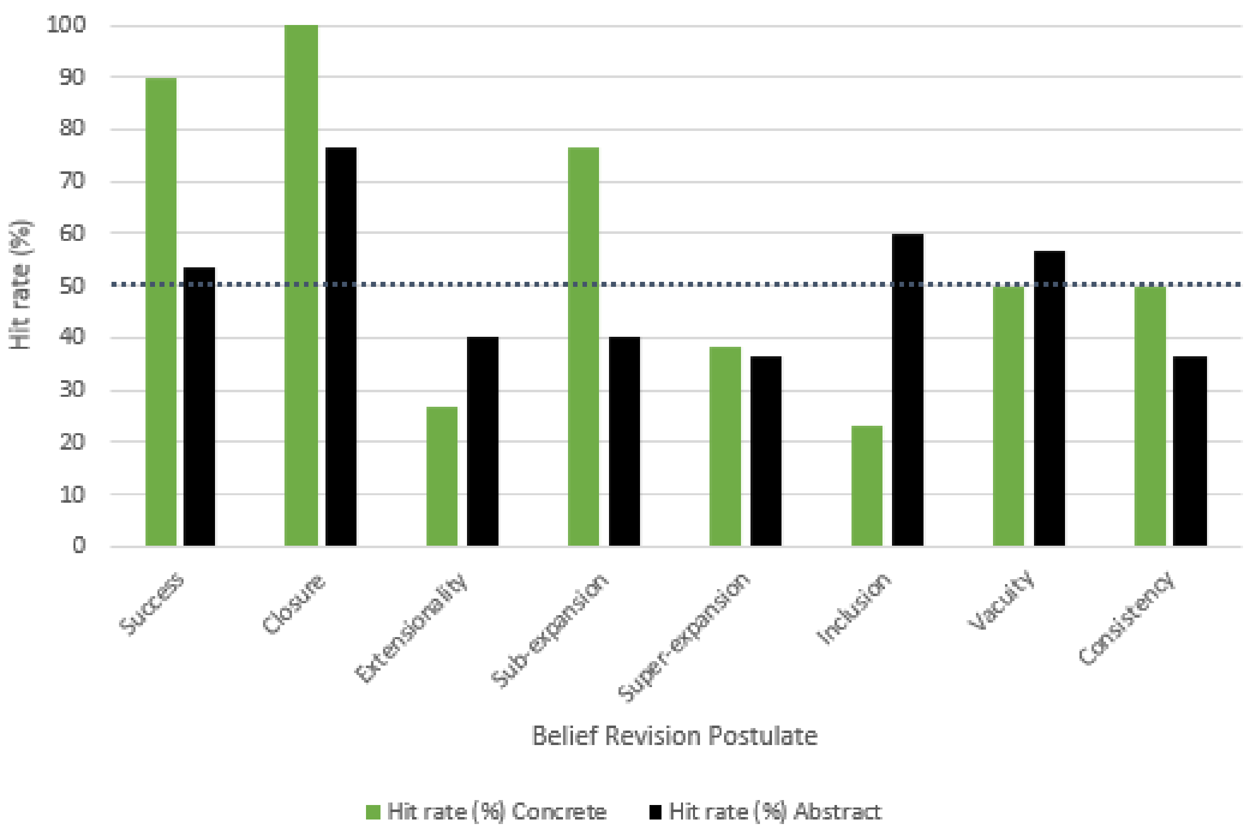

Figure 5: Bar graph showing the Hit Rate per postulate for the corresponding concrete and abstract questions in the final survey. For postulates that have more than one concrete question pertaining to it, the average of the hit rates for those questions are taken. There is a 50% line included for analysis purposes.

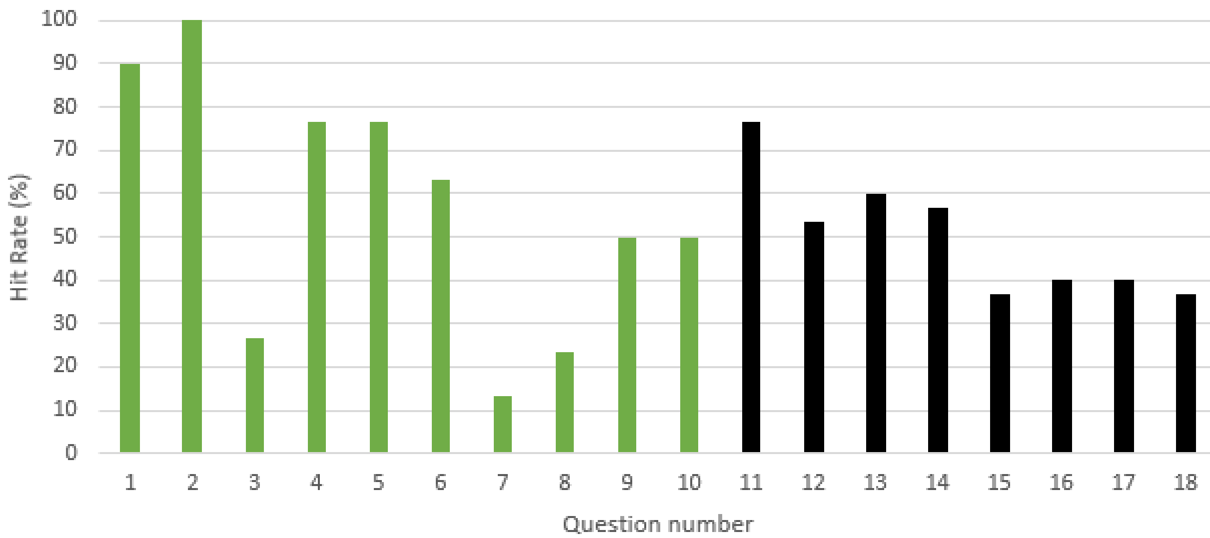

Figure 6: Bar graph showing the Hit Rate per question for the questions in the final survey. Questions 1-10 are concrete and Questions 11-18 are abstract.

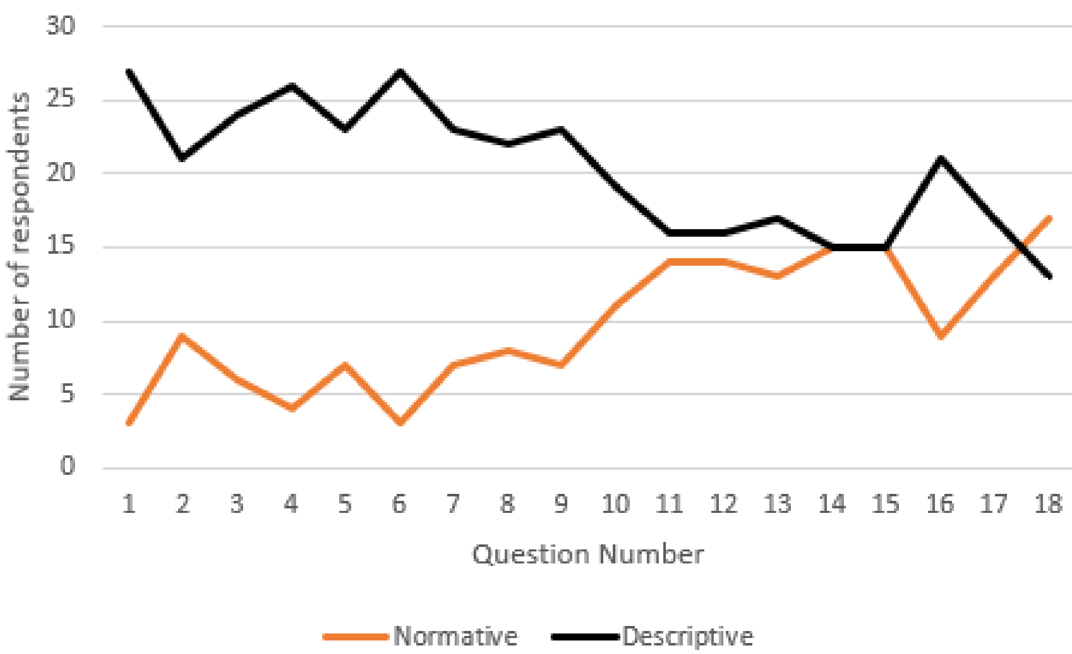

Figure 7: Graph showing the distribution of normative and descriptive reasoning styles for each question. Questions 1-10 are concrete and Questions 11-18 are abstract.

S2 : TABLES

CONCRETE QUESTIONS

Question Number	*Answer*	*Postulate*
1	Yes	Success
2	Yes/No	Closure
3	No	Extensionality
4	Yes, in both case 1 and 2	Sub-expansion
5	Yes, in both case 1 and 2	Sub-expansion
6	Yes, in both case 1 and 2	Super-expansion
7	Yes, in case 2	Super-expansion
8	Yes, in case 2	Inclusion
9	Yes, in both case 1 and 2	Vacuity
10	No	Consistency

Table 1: Table showing survey answers to concrete questions (Q1-Q10), obtained using Belief Revision

ABSTRACT QUESTIONS

Question Number	*Answer*	*Postulate*
11	Yes	Closure
12	Yes, in both case 1 and 2	Success
13	Yes	Inclusion
14	Yes	Vacuity
15	Yes	Consistency
16	No	Extensionality
17	No	Sub-expansion
18	No	Super-expansion

Table 2: Table showing survey answers to abstract questions (Q11-Q18), obtained using Belief Revision

Question Number	Belief Base	New Information	Question
1	$k_1 \wedge k_2$	α	α?
2	$k_1 \wedge k_2$	α	k_2?
3	$k_1 \wedge k_2 \wedge k_3 \wedge k_4$	$\alpha_{case1} = \beta_1 \Rightarrow \beta_2$	α_{case1}?
	$k_1 \wedge k_2 \wedge k_3 \wedge k_4$	$\alpha_{case2} = \beta_1 \nrightarrow \neg\beta_2$	α_{case2}?
4	$k_1 \wedge k_2$	$\alpha_{case1} = \beta_1 \wedge \beta_2$	β_1?
	$(k_1 \wedge k_2 \wedge \beta_1) \vee \beta_2$		β_1?
5	$k_1 \wedge k_2$	$\alpha_{case1} = \beta_1 \wedge \beta_2$	k_1?
	$(k_1 \wedge k_2 \wedge \beta_1) \vee \beta_2$		k_1?
6	$k_1 \wedge k_2$ where $k_1 = x \Rightarrow y$	$\alpha_{case1} = \beta_1 \wedge \beta_2$ where $\beta_1 = z \Rightarrow x$, $\beta_2 = z \Rightarrow \neg y$	z?
	$(k_1 \wedge k_2 \wedge \beta_1) \vee \beta_2$ where $k_1 = x \Rightarrow y$, $\beta_1 = z \Rightarrow x$, $\beta_2 = z \Rightarrow \neg y$		z?
7	$k_1 \wedge k_2$ where $k_1 = x \Rightarrow y$	$\alpha_{case1} = \beta_1 \wedge \beta_2$ where $\beta_1 = z \Rightarrow x$, $\beta_2 = z \Rightarrow \neg y$	k_1?
	$(k_1 \wedge k_2 \wedge \beta_1) \vee \beta_2$ where $k_1 = x \Rightarrow y$, $\beta_1 = z \Rightarrow x$, $\beta_2 = z \Rightarrow \neg y$		k_1?
8	$k_1 \wedge k_2$ where $k_2 = x \Rightarrow (y_1 \wedge y_2 \wedge y_3)$	$\alpha_{case1} = \beta = x \nrightarrow y_3$	β?
	$(k_1 \wedge k_2) \vee \beta$ where $k_2 = x \Rightarrow (y_1 \wedge y_2 \wedge y_3)$, $\beta = x \nrightarrow y_3$		β?
9	$k_1 \wedge k_2$ where $k_2 = x \Rightarrow (y_1 \wedge y_2 \wedge y_3)$	$\alpha_{case1} = \beta = x \Rightarrow (y_1 \wedge y_2 \wedge y_3 \wedge y_4)$	β?
	$(k_1 \wedge k_2) \vee \beta$ where $k_2 = x \Rightarrow (y_1 \wedge y_2 \wedge y_3)$, $\beta = x \Rightarrow (y_1 \wedge y_2 \wedge y_3 \wedge y_4)$		β?
10	$k_1 \wedge k_2 \wedge k_3$	$\alpha = \beta_1 \wedge \beta_2 \wedge \beta_3$ where $\beta_2 = \beta_1 \Rightarrow \neg\beta_1$	β_3?

Table 3: Table giving abstract representations from which the concrete questions were further developed. Each line per question represents a case to which the question refers, for example Question 2 has one case and Question 3 has two cases.

Question Number	Postulate	Normative	Descriptive	Total	Overall classification
1	Success	3	27	30	Descriptive
2	Closure	9	21	30	Descriptive
3	Extensionality	6	24	30	Descriptive
4	Sub-expansion	4	26	30	Descriptive
5	Sub-expansion	7	23	30	Descriptive
6	Super-expansion	3	27	30	Descriptive
7	Super-expansion	7	23	30	Descriptive
8	Inclusion	8	22	30	Descriptive
9	Vacuity	7	23	30	Descriptive
10	Consistency	11	19	30	Descriptive
11	Closure	14	16	30	Descriptive
12	Success	14	16	30	Descriptive
13	Inclusion	13	17	30	Descriptive
14	Vacuity	15	15	30	Other
15	Consistency	15	15	30	Other
16	Extensionality	9	21	30	Descriptive
17	Sub-expansion	13	17	30	Descriptive
18	Super-expansion	17	13	30	Normative

Table 4: Table showing distribution of normative and descriptive reasoning styles for each question. Questions 1-10 are concrete questions and Questions 11-18 are abstract questions.

Reasoning style	Number of questions
Descriptive	15
Normative	1
Other	2

Table 5: Table showing distribution of reasoning style over the questions

Reasoning style	Number of respondents
Descriptive	24
Normative	4
Other	2

Table 6: Table showing distribution of reasoning style over respondents

Question Number	Modal answer	Count	Belief Revision answer	Hit Rate (%)
1	Yes	27	Yes	90
2	Yes	15	Either, explanation-dependent	100
3	Yes, in case 2	10	No	26.67
4	Yes, in both case 1 and 2	23	Yes, in both case 1 and 2	76.67
5	Yes, in both case 1 and 2	23	Yes, in both case 1 and 2	76.67
6	Yes, in both case 1 and 2	19	Yes, in both case 1 and 2	63.33
7	Yes, in both case 1 and 2	17	Yes, in case 2	13.33
8	No	16	Yes, in case 2	23.33
9	Yes, in both case 1 and 2	15	Yes, in both case 1 and 2	50
10	No	15	No	50
11	Yes	23	Yes	76.67
12	Yes, in both case 1 and 2	16	Yes, in both case 1 and 2	53.33
13	Yes	18	Yes	60
14	Yes	17	Yes	56.67
15	No	19	Yes	36.67
16	Yes	18	No	40
17	Yes	18	No	40
18	Yes	19	No	36.67

Table 7: Table showing the modal answer, correct answer and hit rate per question