UNIVERSITY OF CAPE TOWN

DEPARTMENT OF COMPUTER SCIENCE

CS/IT Honours
Final Paper 2019

Title: Evaluating the Katsuno Mendelzon postulates for belief update against human reason

Author: Paul Freund

Project Abbreviation: CDR

Supervisor(s): Professor T. Meyer and Associate Professor D. Moodley

Category	*Min*	*Max*	Chosen
Requirement Analysis and Design	*0*	*20*	
Theoretical Analysis	*0*	*25*	5
Experiment Design and Execution	*0*	*20*	20
System Development and Implementation	*0*	*20*	
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*	
Total marks		**80**	

Evaluating the Katsuno-Mendelzon postulates for belief update against human reasoning

Paul Freund
University of Cape Town
Cape Town, South Africa
frnpau013@myuct.ac.za

ABSTRACT

Classical logic is the default way of modelling idealized human reasoning. Yet it is of limited use when it comes to modelling how to reason in the face of new information. This paper reports on an experiment designed to test the extent to which the postulates proposed by the KM approach to belief update, a non-classical extension of propositional logic, conform to human reasoning. The experiment was split into three sections. The first tested the degree to which participants agreed with a non-formal rendition of the KM postulates, the second tested whether people agreed with a confirming example for each postulate, and the final section tested several purported counter-examples to the postulates. It was found that while for each postulate there exist instances in which participants agreed with the prescriptions of the postulate, participants only found half of the non-formal renditions of the postulates intuitively agreeable, and the counter-examples to the postulates were generally agreed with.

CCS CONCEPTS

• **Theory of computation → Semantics and reasoning**; *Logic.*

KEYWORDS

propositional logic, belief change, belief update

1 INTRODUCTION

Reasoning is an integral part of human life. It is well-documented that people fail to conform to the prescriptions of classical logic [27] [26], the paradigm model of reasoning in mathematics and computer science. These deviations are thought to be systemic [27], in that they are not arbitrary, as many people will make the same non-classical inference on identical problems. It is thought that these deviations model a flexibility not found in classical logic [24]. The artificial intelligence community therefore seeks to incorporate this flexibility in its work [24].

On this basis various extensions to propositional logic have been proposed in the artificial intelligence community, such as the AGM approach to belief revision [1], the KLM approach to defeasible reasoning [20], and the KM approach to belief update [18]. In these approaches a set of postulates governing the non-classical part of the logic is proposed. The postulates governing the approaches are presented as reasonable on an *a priori* basis; what seems reasonable to the academic in the armchair. While the postulates may be considered normative in nature, i.e. what idealized human reasoning should be in the respective domains, given that the purported relevance of the topics is often based in *actual* human reasoning, there seems to be space for testing empirically whether the postulates accurately model human reasoning, even without adopting full-blown psychologism (that there is no standard apart from that of everyday human practice) in these domains, such as in [24]. So far research in this area has been concerned with testing whether human reasoning follows the postulates of various systems of defeasible reasoning [27] [26] [24]. This project is concerned with extending this emerging research paradigm to the case of belief change, or the question as to agents change their beliefs in response to new information. In particular, it is concerned with testing the degree to which people reason in conformance with the KM approach [18] to belief update.

This requirement has been deconstructed into three components, corresponding to three research questions:

R1: How intuitively plausible do people find the KM postulates?
R2: For each of the KM postulates, are there instances in which people reason in accordance with them?
R3: Are there any instances in which people do not reason in accordance with the KM postulates?

To test the research questions an experiment was conducted consisting of a survey posted on Amazon's Mechanical Turk [23], an online portal for crowdsourced work. The survey is split into four sections, three of which correspond to the three research questions, and the final section being a section to test the participants' perception of the quality of the survey.

2 BACKGROUND

2.1 Propositional Logic

Let $\mathcal{P}$ be a non-empty set, and $p, q, \ldots$ refer to members of $\mathcal{P}$, henceforth atoms. Let $\mathcal{L}$ be the set of all α generated by $\alpha := \top \mid \bot \mid p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \alpha \leftrightarrow \alpha$. $\mathcal{L}$ is the language of propositional logic, and its members are propositional sentences, denoted by $\alpha, \beta, \ldots$. Let $\mathcal{U}$ refer to the set of all interpretations $u : \mathcal{P} \mapsto \{0, 1\}$, denoted by $u, v, \ldots$. Intuitively, if an atom is assigned 1 under an interpretation this means that under that interpretation that atom is true, and if it is assigned 0, then it is false.

Given an interpretation, the interpretation is extended to all sentences of $\mathcal{L}$ by the valuation functional $V : \mathcal{U} \mapsto (\mathcal{L} \mapsto \{0, 1\})$ in the usual manner as outlined in [3]. We say $u \in \mathcal{U}$ satisfies $\alpha \in \mathcal{L}$ if and only if $V_u(\alpha) = 1$, and if such a u exists then we say α is *satisfiable.* The set of interpretations that satisfy a sentence, α, is the set of models for that sentence, denoted as $Mod(\alpha)$. The models of a set of sentences, $Mod(\mathcal{K})$, is the set of interpretations such that for each $\alpha \in \mathcal{K}$, $V_u = 1$. Occasionally it shall be necessary to refer to the power set of a given set, V. This shall be denoted by $P(V)$. Finally, $a \oplus b$ is used as an abbreviation for $(a \wedge \neg b) \vee (\neg a \vee b)$.

2.2 Entailment

Consider now a set of sentences, $\mathcal{K}$, which represents what an agent knows about the world. Such a $\mathcal{K}$ is dubbed a knowledge base. Given a particular $\mathcal{K}$, intuitively we may reason from the sentences of $\mathcal{K}$ to new sentences, and thus come to know more about the world.

For example, take $\mathcal{K} = \{p, p \rightarrow q\}$, where p represents the proposition that Socrates is human, and $p \rightarrow q$ represents the proposition that if Socrates is human then Socrates is mortal (the classical reading of $\alpha \rightarrow \beta$ is "if α then β"). Then we may infer q, that Socrates is mortal. By way of contrast, few would regard the inference that the moon is made of cheese as warranted given this $\mathcal{K}$. Whether a particular proposition follows from a knowledge base is dependent both on the contents of that knowledge base, and the meaning of that sentence.

So consider a relation $\models \subseteq P(\mathcal{L}) \bigtimes \mathcal{L}$, where $\mathcal{K} \models \alpha$ is to be interpreted as $\mathcal{K}$ entails α. It was Tarski and Gentzen (in e.g. [30] and [13]) who popularized the idea that by studying the properties of such a relation, we may thereby come to better understand the deductive process.

One such property is truth-preservation; that $\mathcal{K} \models \alpha$ may hold if and only if whenever all the sentences in $\mathcal{K}$ are all true, that α be true as well. Phrased semantically, this is equivalent to $\mathcal{K} \models \alpha$ if and only if $Mod(\mathcal{K}) \subseteq Mod(\alpha)$. Clearly such a property uniquely defines a single $\models$ and this is classically considered the only notion of entailment. Thus, henceforth $\models$ shall refer solely to that classical notion. The consequence set of a knowledge base is denoted $Cn(\mathcal{K})$ and is equal to $\{\gamma : \mathcal{K} \models \gamma\}$.

2.3 Belief Revision

The classical account of entailment answers the question as to what it would be reasonable to believe given a fixed set of beliefs. Yet it fails to accurately model many aspects of our actual reasoning practices. We often come into contact with new information, such that this new information is not entailed by what we previously believe. Furthermore this information may contradict some of what we already believe. In which case, if we were to use the classical account of entailment to model what it would now be reasonable to believe, we would find that it would be reasonable to believe anything. This is as, as a contradiction has no models, every model of a contradiction is a model of any sentence.

This is highly counter-intuitive. In our everyday lives we often come across information that contradicts what we previously believed, and it is certainly not generally believed that it is reasonable to believe anything. So if we are to model this aspect of our inferential practices then an alternative to the classical account of entailment is needed.

The central insight of the AGM approach [1] to studying such dynamics, is that they may be modelled by a binary operator $*$ between an already extent theory (a deductively closed set) which an agent believes, and a new fact, where the output of that binary operator is *another* theory. The intended interpretation is that we start with all that may be concluded, model the finding of new information by the binary operation, and by studying the properties of the operator we may thereby better come to understand what may be concluded given this new information, for that is the result of the operation.

While it is well-accepted that, unlike in the classical case, there need not be a *unique* such operator, the focus is on the properties that *any* such acceptable $*$ need fulfill. The AGM approach proposes a set of properties (not presented here), where the guiding notion is that $*$ must respect the so-called *principle of minimal change* or *principal of the informational economy*, as presented in [28]:

- When an agent learns α, she should adopt a posterior belief set, $\mathcal{T}'$, such that (i) $\mathcal{T}'$ is deductively cogent, (ii) $\mathcal{T}'$ includes α, and (iii) $\mathcal{T}'$ is the closest belief set to her prior belief set $\mathcal{T}'$, which satisfies (i) and (ii).

2.4 Belief Update

In the context of belief revision we are concerned with modelling the dynamics of reasoning upon the learning of new information. It was in the database community [19] [31] that it was first noticed that depending on the context in which that new information is learned, different dynamics of reasoning may be appropriate, even for formally identical examples [18]. In particular a distinction is drawn between learning new information about an unchanging world, and learning of the world that it has undergone new changes. The former is what is associated with belief revision, and the latter is dubbed that of belief update.

To get an intuitive grasp of the distinction between belief update and revision, take the following example adapted from [18]. Let b be the proposition that the book is on the table, and m be the proposition that the magazine is on the table. Say that our belief set, $\mathcal{T}$, may be represented by the deductive closure of $(b \wedge \neg m) \vee (\neg b \wedge m)$, that is the book is on the table or the magazine is on the table, but not both. We send a student in to report on the state of the book. She comes back and tells us that the book is on the table, that is b. Intuitively, given that we initially believed that at most one of the book and the magazine were on the table, it would now be reasonable to believe that the magazine is not on the table. And if a believe revision operator obeyed the postulates proposed by the AGM approach, $\neg m$ would indeed be a consequence.

But consider now that instead of asking the student to report on the state of the book we had instead asked her to ensure that the book were on the table. After reporting back that she had indeed ensured that the book is on the table, we again are faced with the new knowledge that b. Now it seems presumptuous to assume that the magazine is not on the table [18]. Either the book were already on the table and the magazine were not, in which case the student would have done nothing and left, or the magazine were on the table and the book not, in which case the student presumably would have simply put the book on the table and left the magazine similarly so on.

It is on the basis of such examples that Katsuno and Mendelzon [18] argue that within the context of belief update, what they dub the *disjunction rule* is appropriate: the result of updating $\gamma \vee \phi$ with α is equivalent to the disjunction of γ updated with α and ϕ updated with α. The disjunction rule is inconsistent with the AGM postulates. So they proposed some new properties, ones more appropriate to belief update.

Table 1: KM Postulates

(U1)	$\phi \diamond \alpha \models \alpha$
(U2)	If $\phi \models \alpha$ then $\phi \diamond \alpha = \phi$
(U3)	If both ϕ and α are satisfiable then $\phi \diamond \alpha$ is satisfiable
(U4)	If $\phi_1 \leftrightarrow \phi_2$ and $\alpha_1 \leftrightarrow \alpha_2$ then $\phi_1 \diamond \alpha_1 \leftrightarrow \phi_2 \diamond \alpha_2$
(U5)	$(\phi \diamond \alpha) \wedge \gamma \models \phi \diamond (\alpha \wedge \gamma)$
(U6)	If $\phi \diamond \alpha_1 \models \alpha_2$ and $\phi \diamond \alpha_2 \models \alpha_1$ then $\phi \diamond \alpha_1 \leftrightarrow \phi \diamond \alpha_2$
(U7)	If ϕ is complete then $(\phi \diamond \alpha_1) \wedge (\phi \diamond \alpha_2) \models \phi \diamond (\alpha_1 \vee \alpha_2)$
(U8)	$(\phi_1 \vee \phi_2) \diamond \alpha \leftrightarrow (\phi_1 \diamond \alpha) \vee (\phi_2 \diamond \alpha)$

These properties are summarized in Table 1. Again, they do not single out a unique operator, rather they specify a family of admissible operators. To understand them it first has to be understood that the formal context in which Katsuno and Mendelzon are working is an adaption of the AGM approach [17]. Instead of working directly with deductively closed sets, $\mathcal{T}$, they work only with such sets that have a finite cover, that is a finite set of sentences $\mathcal{K}$ such that $Cn(\mathcal{K}) = \mathcal{T}$, and instead of working directly with that $\mathcal{K}$ they work with a representational sentence $\phi \in \mathcal{L}$ such that $\mathcal{K} = \{\gamma : \{\phi\} \models \gamma\}$. So the binary operator they are working with is a sentential operator, it takes in sentences of propositional logic and outputs a new sentence, with the new sentence standing as representative of the new knowledge base. What follows is an explanation of the postulates in ordinary language:

(U1): states that updating with the new fact must ensure that the new fact is a consequence of the update.
(U2): states that updating on a fact that could in principle be already known has no effect.
(U3): states the reasonable requirement that we cannot lapse into impossibility unless we either start with it, or are directly confronted by it.
(U4): requires that syntax is irrelevant to the results of an update.
(U5): first updating on α then simply adding the new information γ is at least as strong (i.e. entails) as updating on the conjunction of α and γ.
(U6): states that if updating on α_1 entails α_2 and if updating on α_2 entails α_1, then the effect of updating on either is equivalent.
(U7): applies only to complete ϕ, that is ϕ which have only one model. If some situation arises from updating a complete ϕ on α_1 and it also results from updating that ϕ from α_2 then it must also arise from updating that ϕ on $\alpha_1 \vee \alpha_2$ [18].
(U8): The aforementioned disjunction rule.

Katsuno and Mendelzon also provide a semantic characterization of the behaviour of admissible update operators. First, take a function that associates with each interpretation u a partial preorder (a reflexive, transitive relation that may not be total), $\leqslant_u$ on $\mathcal{U} \times \mathcal{U}$. Such an assignment is called faithful if and only if for all $v \in \mathcal{U}$, if $v \neq u$ then $u \leqslant_u v$ and it is not the case that $v \leqslant_u u$; i.e. u is a minimum element of $\mathcal{U}$ under $\leqslant_u$. Next define a function $Min : P(\mathcal{U}) \times \mathcal{U} \mapsto P(\mathcal{U})$ such that $Min(V, u)$ is equal to the set of elements in V that are minimal under $\leqslant_u$ for the set V. Finally, we have the following result:

Theorem 2.1 ([18]). *Let $\diamond$ be an update operator. The following conditions are equivalent:*

- *$\diamond$ satisfies postulates **U1 - U8***
- *There exists a faithful assignment that assigns each interpretation u a partial pre-order $\leqslant_u$ such that $Mod(\phi \diamond \alpha) = \bigcup_{u \in Mod(\phi)} Min(Mod(\alpha), u)$.*

3 OBJECTIONS TO THE KM POSTULATES

While the AGM postulates for belief revision have been generally accepted in the literature, the KM postulates for update have seen less acceptance. In this section four objections to the KM approach to belief revision are discussed.

The KM approach to belief update was introduced as a generalization of Winslett's Possible Model Approach (PMA) [31]. Define the difference between two interpretations, $D : \mathcal{U} \times \mathcal{U} \mapsto P(\mathcal{P})$, as the set of atoms that are assigned different values under those interpretations. In the PMA the faithful preorders needed for the semantic characterization of update are generated by the following: $v \leqslant_u w$ if and only if $D(v, u) \subseteq D(w, u)$.

A common criticism of the PMA is that it handles disjunctive input counter-intuitively (see e.g. [11], [6]). This comes in two forms. Using conjunction of literals as specifications of interpretations, consider the following two examples:

Example 1. Let $\mathcal{P} = \{p, q\}$, $\phi = p$ and $\alpha = p \vee q$. So $Mod(\phi) = \{p \wedge q, p \wedge \neg q\}$ and $Mod(\alpha) = \{p \wedge q, p \wedge \neg q, \neg p \wedge q\}$. $Min(Mod(\alpha), p \wedge q) = \{p \wedge q\}$ and $Min(Mod(\alpha), p \wedge \neg q) = \{p \wedge \neg q\}$ so $Mod(\phi \diamond \alpha) = Mod(\phi)$ so $\phi \diamond \alpha$ may be represented by p.

Example 2. Let $\mathcal{P} = \{p, q\}$, $\phi = \neg p \wedge \neg q$ and $\alpha = p \vee q$. $Mod(\phi) = \{\neg p \wedge \neg q\}$. and $Mod(\alpha) = \{p \wedge q, p \wedge \neg q, \neg p \wedge q\}$. Say that u is the interpretation specified by $\neg p \wedge \neg q$. For the interpretations in $Mod(\alpha)$ we have that $p \wedge \neg q \leqslant_u p \wedge q$ and $\neg p \wedge q \leqslant_u p \wedge q$ and no more. So $Min(Mod(\alpha), u) = \{p \wedge \neg q, \neg p \wedge q\} = Mod(\phi \diamond \alpha)$ so ϕ may be represented by $p \oplus q$.

Example 1 demonstrates that in the PMA if p is already believed then updating by $p \vee q$ has no effect. Now consider the case that the atoms p and q are interpreted as in the last section, i.e. p is the proposition that a book is on the table, and q is the proposition that a magazine is on the table. Say we initially believe p and that we then ask a robot to ensure that $p \vee q$ is the case.

As noted in [6] the result that the interpretation specified by $\neg p \wedge q$ is not among the believed possible models of the updated world only makes sense if we additionally assume both *awareness* and *laziness*: that the robot becomes aware during its task that $p \vee q$ is already satisfied, and that if it knows $p \vee q$ is already satisfied, then it does not attempt to bring about any alternative state of satisfaction. And in the case of lacking further information about the robot, this seems like perhaps too much to assume. As U2 ensures that $p \diamond (p \vee q) = p$ this criticism of the PMA carries over to the KM approach to belief update.

Example 2 demonstrates that in the PMA that if $\neg p \wedge \neg q$ is initially believed then updating by the inclusive disjunction has it that the exclusive disjunction is believed. This is problematic for the following example [15]: suppose you are throwing a coin on a chessboard. Let p be the proposition that the coin is on a white field, and q be the proposition that the coin is on a black field. Initially you are holding the coin so you believe $\neg p \wedge \neg q$. After the coin falls on the chessboard you are certain that $p \vee q$ holds, that the coin is either on a white field or on a black field. But why should it be believed that it is not possible that the coin is *both* on a white field or on a black field?

Unlike in the case of Example 1 it is not immediately obvious that this criticism of the PMA carries over to the generalized setting of the KM approach to belief update. The problem, of course, is that there are no requirements on the faithful preorder other than that it be faithful, so in the general case for the example there is nothing prohibiting e.g. $p \wedge q \leqslant_u p \wedge \neg q$.

Fortunately we have the following result, which while not exactly analogous, is in the same spirit:

Lemma 3.1 ([16]). *Say an update operator, $\diamond$, satisfies U1, U4, and U5. Then if $\phi \diamond \alpha \models \neg\beta$ and $\phi \diamond \beta \models \neg\alpha$ then $\phi \diamond (\alpha \vee \beta) \models \alpha \oplus \beta$.*

Proof. Suppose $\phi\diamond\alpha \models \neg\beta$ and $\phi\diamond\beta \models \neg\alpha$. By U1 $\phi\diamond(\alpha\vee\beta) \models (\phi\diamond(\alpha\vee\beta))\wedge(\alpha\vee\beta) \models (\phi\diamond(\alpha\vee\beta))\wedge\alpha\vee(\phi\diamond(\alpha\vee\beta))\wedge\beta = \gamma$. By U5 $\gamma \models (\phi\diamond((\alpha\vee\beta)\wedge\alpha))\vee(\phi\diamond((\alpha\vee\beta)\wedge\beta))$, so by U4, $\gamma \models (\phi\diamond\alpha)\vee(\phi\diamond\beta)$. By hypothesis and U1 $(\phi\diamond\alpha)\vee(\phi\diamond\beta) \models (\alpha\wedge\neg\beta)\vee(\neg\alpha\wedge\beta)$ which is equivalent to $\alpha\oplus\beta$. So $\phi\diamond(\alpha\vee\beta) \models \alpha\oplus\beta$. □

Contra Herzig and Rifi [16], while this result may point to a similar problem for the KM approach to belief update, the present author does not think that the coin example in particular poses a problem - if we were told that the coin had fallen on a black field then there does not seem to be any reason to believe that it had not fallen on a white field. But say that we initially believe that $\neg p \wedge \neg q$ and p is in some (non-formal) manner independent of q. Then intuitively $\neg p \wedge \neg q \diamond p \models \neg q$ and $\neg p \wedge \neg q \diamond q \models \neg p$, but $\neg p \wedge \neg q \diamond (p \vee q) \models p \oplus q$ is counter-intuitive - there seems to be no reason to assume that $p \wedge q$ could not be the case (for a concrete example, again try interpreting p and q and the update operation as in the robot example).

The next objection to the KM approach is not (directly) to the postulates themselves, but rather as to how belief update is conceived. The difference between belief update and belief revision is commonly held to be, and has been presented here as, the difference between learning new information of a static world, versus learning of a dynamic world that it has undergone new changes.

Yet the difference between the two does not seem to be that simple. If we consider a time-indexed language, that is, each statement is accompanied by a time that specifies when it is occurring then the AGM model of belief revision seem to accommodate the changes in the world just as well as the KM approach to belief update [12]. It is such considerations that have led some to conceive as what is crucial for belief revision as opposed to belief update is not that the *world* is unchanging, but rather that the language we use to describe it is [12].

However, as noted by Lang [22], there are cases in which the language we use is static, what we initially believe is true, and the world changes, but belief revision seems more appropriate than belief update. Consider the following example:

Example 3. *You work in an office with Alice and Bob. Both tend to stay in the office when they are in, and stay out of the office when they are out. You initially believe that at most one of Alice and Bob are in the office. You then see Bob leave the office. What do you now believe?*

Let a be the proposition that Alice is in the office and b be the proposition that Bob is in the office. Then initially you believe $\phi = a \oplus b$ which is equivalent to $(a \wedge \neg b) \vee (\neg a \wedge b)$. When Bob leaves the office we have it that $\neg b$. By U8 $\phi \diamond \neg b$ is equivalent to $(a\wedge\neg b)\diamond\neg b \vee (\neg a\wedge b)\diamond\neg b$. Intuitively, (but not by the KM postulates which are too weak to derive this) as a and b are independent $(a \wedge \neg b) \diamond \neg b = (a \wedge \neg b)$ and $(\neg a \wedge b) \diamond \neg b = (\neg a \wedge \neg b)$ so that $\phi \diamond \neg b$ is equivalent to $\neg b$. But given that we initially believed that at most one of Alice and Bob were in the office and then saw Bob leave, it seems that we can conclude *more* than this in this case, namely that Alice is not in the office. The point of the example is that sometimes new information about the world *can itself* serve as evidence to invalidate one of our previous possible models of the world, without making what we previously believed of the world false. In such cases U8, the hallmark of belief update, seems no longer appropriate.

Lang [22] does not see this as a counter-example to the KM postulates themselves. Instead, he sees the example as necessitating a change in our conception of where belief update is appropriate. In an experimental setting however, such a move is problematic. This is as if what one is testing is contested, then what one is testing is unclear, (unless what one is testing is various conceptions of belief update, which this project is not doing). Rather, it is open to take belief update as it is has been presented previously, see the example as providing a counter-example to U8, and the *import* of the example (if it is found that people agree with it) being that a reconceptualization of where belief update is appropriate may be necessary.

The final objection to the KM postulates is the author's own. It is a combination of the intuition that $(\neg p \wedge \neg q) \diamond p \models \neg q$ in the case that p is independent of q (which was used in the example that uses Lemma 3.1), and the intuition active in the example against $U2$ that when it comes to disjunctive input, each model of that input deserves equal consideration in the final output. Putting these two together we have that p is equivalent to $(p \wedge q) \vee (p \wedge \neg q)$. So by U4, which implies that equivalent updates leads to equivalent results, $(\neg p \wedge \neg q) \diamond p$ is equivalent to $(\neg p \wedge \neg q) \diamond ((p \wedge q) \vee (p \wedge \neg q))$. But in the case that p is independent of q (again, for a concrete example try interpreting as in the robot example) it seems to the author that $(\neg p \wedge \neg q) \diamond p \models \neg q$ but this does not hold for $(\neg p \wedge \neg q) \diamond ((p \wedge q) \vee (p \wedge \neg q))$, so the two are not equivalent.

As a final note, per [16], U4 may be decomposed into the following, as any sentence is equivalent to itself:

(U4.1) If $\phi_1 \leftrightarrow \phi_2$ then $\phi_1 \diamond \alpha \leftrightarrow \phi_2 \diamond \alpha$
(U4.2) If $\alpha_1 \leftrightarrow \alpha_2$ then $\phi \diamond \alpha_1 \leftrightarrow \phi \diamond \alpha_2$

Lemma 3.2. *Let $\diamond$ be an update operator. $\diamond$ satisfies U1 and U6 only if it satisfies U4.2.*

Proof. Say $\alpha_1 \leftrightarrow \alpha_2$. By U1 $\phi \diamond \alpha_1 \models \alpha_1$ so by hypothesis $\phi \diamond \alpha_1 \models \alpha_2$. Similarly $\phi \diamond \alpha_2 \models \alpha_2 \models \alpha_1$. So by U6, $\phi \diamond \alpha_1 \leftrightarrow \phi \diamond \alpha_2$. □

The just discussed example follows from U4.2 so it also follows from any update operator that obeys U1 and U6. It thus also serves as a counter-example to either U1 or U6, or both.

4 EXPERIMENT DESIGN AND EXECUTION

4.1 Survey Overview

4.1.1 Section 1.

In the first section of the survey the KM postulates are presented and people are to rate their agreement with the postulates on a linear or Likert scale (five options with extremal points 'strongly disagree' and 'strongly agree'). The postulates are presented using non-technical language (for example, using the word 'possible' as a proxy for 'satisfiable') so that the lay-person could understand them. This section of the survey is intended to test research question 1. The major design question here was whether to present participants with a yes/no response or to present participants with a Likert scale response. Initially the plan was to use yes/no responses and to have participants type out a reason for their answer, so that the responses would have both a quantitative and qualitative component. However, early on in designing the survey it became apparent that keeping the survey down to a manageable length to participants was going to be an issue. As a result, it was decided that the reason for the response by the participants needed to be cut. The change to a Likert scale type response allows for finer-grade differentiation of opinion as compared to a simple yes/no response, corresponding to the degrees of confidence participants feel in their affirmation or negation of the KM postulates.

4.1.2 Section 2.

In the next section participants are presented with examples that are meant to be confirming instances of the KM postulates. This section is intended to test research question 2. Table 2 on the next page outlines the questions asked in this section. The examples are mostly presented in symbolic form in the table; i.e. in the language of propositional logic. However, this is purely for brevity - in the survey proper the questions are phrased in ordinary English, and use actual propositions instead of symbols that represent them.

Each question must be answered with a yes/no response. In Table 2 for each question the expected answer for that question is indicated in brackets. Additionally, in the table, in the New Information column there may be multiple numbered items. Updating the initial beliefs by the information numbered corresponds to the different situations that are referred to in the Questions column.

As for how the questions test the postulates, this should be self-explanatory, except for the the questions testing U7 and U5. For U7 note that if the answer is yes to both questions then $((a \wedge b) \diamond \neg a) \wedge ((a \wedge b) \diamond \neg b)$ is inconsistent, in which case U7 trivially follows. For U5 note that if the answer is yes to both the initial questions, then $(a \oplus b) \diamond b$ is equivalent to b for the domain in question, so $((a \oplus b) \diamond b) \wedge a$ is equivalent to $a \wedge b$. The final question then asks if $(a \oplus b) \diamond (a \wedge b)$ models (and thus is equivalent for the domain in question) $a \wedge b$. So if the answer is yes then $((a \oplus b) \diamond b) \wedge a \models (a \oplus b) \diamond (a \wedge b)$, as would be expected given U5.

Participants must also provide a reason for their choice. This is so as to better understand any patterns of reasoning that emerge, and whether such patterns correspond with the KM postulates.

4.1.3 Section 3.

The format of Section 3, and the types of responses, is the same as that of Section 2. However, this section is meant to provide counter-examples to the KM postulates. In total, four counter-examples are tested, corresponding to the four objections to the KM postulates outlined in the previous section. Again, instead of using propositional logic as was used in the section, concrete examples were chosen to model the situations. The major difficulty here was that, as noted in the section, that often the objections to the postulates rely on intuitions that the KM postulates themselves are too weak to derive. For example, the intuition that $(\neg p \wedge \neg q) \diamond p \models \neg q$ in the case that p and q are independent is not supported in the general case by the KM postulates; i.e. there exist update operators that satisfy the postulates but do not satisfy $(\neg p \wedge \neg q) \diamond p \models \neg q$. To get around this difficulty it was noted that for the examples in question we can test the intuitions of participants by asking them in the survey whether they agree that e.g. $(\neg p \wedge \neg q) \diamond p \models \neg q$ for the example in question. While this solves the problem, it did unfortunately lead to an extending of length of the section, making it nine questions in total, which ultimately necessitated that the reason for participants' responses be cut in Section 1.

4.1.4 Section 4.

This section is an evaluation section for the survey. It consists of six questions. The first asks what the interest of the person taking the survey is, with the options being wanting money, interested in topic, or a student. The participants may choose more than one response. The next four questions are rated on a Likert scale (five options, Strongly Disagree to Strongly Agree). The second question asks the participants to rate the clarity of the instructions. The third asks participants to rate the similarity between the example scenarios presented and situations they would come across in everyday life. The fourth question asks participants to self-report whether they think they reasoned in a manner similar to how they would reason in everyday life. The fifth question asks participants to rate how understandable they considered the questions in section 1. The final question is a long-form question where participants could give any more final thoughts they had on the survey.

4.2 Survey Implementation

The described survey was implemented using Google forms. Initially, a prototype survey was developed.

4.2.1 Gathering of initial feedback.

Before sending the initial survey out for testing, feedback was asked for by five laypeople and an expert (our supervisor) in the field.

For the laypeople who completed the survey it were asked that there be a focus on finding parts of the survey in which meanings were ambiguous, and where instructions were unclear, and highlight them in feedback to the author. That these are desirable aspects of surveys is in conformance with the conventional wisdom about questionnaire design as outlined in [21]. Additionally as the target time of the survey was 30-40 minutes, and there were

Table 2: Section 2 Questions

Initial beliefs	New Information	Question(s)	Postulate(s)
$a \wedge b$	b	Do you believe anything different about the domain? (No)	U2
$a \wedge b$	$\neg a \wedge \neg b$	Would you have to believe anything impossible? (No)	U3
1. $a \vee b$ 2. $\neg(\neg a \wedge \neg b)$	1. $\neg(\neg a \vee \neg b)$ 2. $a \wedge b$	In both situations 1. and 2. do you believe the same things? (Yes)	U4
Lights a and b share the same switch and both lights are off.	1. Light a is on 2. Light b is on	1. In situation 1 do you believe light b is on? (Yes) 2. In situation 2 do you believe light a is on? (Yes) 3. In both situations, do you believe the same things? (Yes)	U6
$a \oplus b$ (robot example)	1. b 2. $a \wedge b$	1. In situation 1 do you believe it is possible that a? (Yes) 2. In situation 1. do you believe b? (Yes) 3. In situation 2. do you believe $a \wedge b$? (Yes)	U1, U5 and U8
$a \wedge b$	1. ¬ a 2. ¬ b	1. In situation 1 do you believe b? (Yes) 2. In situation 2 do you believe $\neg b$? (Yes)	U7

concerns that the survey was overly long, it was also asked that the laypeople evaluating report on the approximate time it took them to complete the survey. For the expert it were asked whether the survey were capable of meeting the research questions.

Based on the initial round of feedback, the survey underwent a rewrite to remove multiple issues regarding ambiguity in questions. All those who completed the survey reported that they did so in under 40 minutes, and thus based on the feedback from this stage it were decided that the survey was of a length that was acceptable for the project. From the expert's side, it were communicated that the proposed survey was fine given the research questions.

4.2.2 Pilot Experiment.

In order to demonstrate the feasibility of the final experiment, a pilot experiment was conducted. This was conducted using Amazon's Mechanical Turk, an online work place where workers can complete tasks for renumeration. The major reason for using Mechanical Turk is that the demographics of Mechanical Turk are thought to be more diverse in both age and geographic spread than that of typical university populations [8]. So Mechanical Turk may yield a more representative sample of the global population than surveys conducted at university, although it should be noted that the majority of Mechanical Turk workers are based in the United States. Additionally, it has been found that respondents in psychological experiments using Mechanical Turk behave similarly to traditional samples, in that many experiments using traditional samples have been successfully replicated using participants sourced from Mechanical Turk [10] [4]. This gives credence to the methodological soundness of using Mechanical Turk as a way of sourcing participants.

The pilot experiment was conducted over a single day, in which the desired five responses were received. Each participant was paid the dollar equivalent of R30 for completing the survey. Additionally, it was required that participants have a 80 percent or greater approval rate for completion of tasks. Requiring such approval ratings is a common practice when using Mechanical Turk for research [9].

The pilot experiment had mixed results. Total response time was relatively fast, in that it only took a day. In addition, mean completion time for the task was just less than thirty minutes. However, of the five responses, for the written responses that needed to be given by workers, three workers gave responses that were essentially incomprehensible.

4.2.3 Final Experiment.

The final experiment ran for a period of three days in early August. Due to the poor quality of responses in the pilot experiment, several changes were made to ensure better quality for the final experiment. First, an attention checker/bot detection section was included in the beginning of the survey. This comprised two items: an attention check whereby participants had to enter a counter-intuitive answer to a question based on text in a previous paragraph, and a question where participants had to enter some randomly generated text. The latter was implemented using Google App Scripts, using a timer such that every five minutes the form is edited with newly randomly generated text. Participants who failed this section were excluded from the final results.

The second step was increasing the compensation to the dollar equivalent of R60, instead of R30, and requiring Masters status of workers as a prerequisite for completing the task. Masters status is awarded to Amazon Turk workers based on excellence on a wide range of tasks, and is seen as a more demanding requirement than requiring than that of approval rating [5].

4.3 Ethical Clearance

As this project involved experiments with people, before proceeding with the experiments ethical clearance was obtained from Faculty of Science Human Research Ethics Committee. The major ethical issue in the project is the use of Mechanical Turk, and in particular whether workers were being paid a fair wage for their work - some have compared Mechanical Turk to an online sweatshop [25]. Per [7] the three steps were taken to mitigate these concerns. First, workers were paid double the South African minimum wage for an hour's work. Second, in the title of the task the estimated amount of time needed for the task was clearly stated. Finally, there is a section in the survey which gives an overview of what the research is concerned with, placing the work in context. Additionally, workers were required to give their informed consent to participate in the study. This was achieved by having a consent form at the start of the survey, whereby workers could either agree to participate in the research and then continue to the rest of the survey, or they could decline to participate and be thanked for their time.

5 RESULTS AND DISCUSSION

5.1 Section 1

Table 3 on the next page shows the results for Section 1 of the survey. It is split into four sections. The first indicates the postulate in question, the second shows the distribution of results, and the third shows the combined disagreement (sum of strongly disagree and disagree) and combined agreement (sum of agree and strongly agree). As the data is ordinal, it is generally agreed that the mean is an inappropriate measure of central tendency [29]. So the median has been chosen as the measure of central tendency. Bolded items in the second section of the table indicate the median. As there were an even number of observations, where the median would fall between two items, both are bolded in the table. Finally, the bottom row of the table sums the results from each postulate to receive an aggregate measure of the postulates as a grouping.

The results indicate that as a whole the postulates were broadly agreed with by participants, with a median value of Agree for the postulates as a grouping, and a 57% combined agreement score as opposed to a 29% combined disagreement score. For the individual postulates, U1, U3, U4, and U6 received the most support from participants, each with a median of Agree and thus combined agreement of over 50%. U2 and U5 received less support, with a median falling between Neutral and Agree. However it should be noted that combined disagreement with U5 was double that of U2. So it seems that the participants saw U2 as less unintuitive than U5. Participants were divided as to their opinion of U8, with a median of neutral and an almost equal split between combined agreement and disagreement. Finally, U7 was the only postulate to receive a negative median (between disagree and neutral), with 50% of participants disagreeing with it and only 36.7% of participants agreeing with it.

Looking to interpret the results, first it should be noted that participants' negative appraisal of U7 is consistent with Herzig and Rifi's contention that U7 is "almost meaningless [16]." And indeed of the seven update operators examined in that paper, only four satisfied U7.

More concerning is the respondents' split opinion of U8, the hallmark of belief update. The exact form of the question is reproduced here:

Example 4. Preamble: *Consider the following three situations. In Situation 1 you initially believe that either P or Q are true, and possibly both. You then learn that R has occurred. In Situation 2 you initially believe that P is true, and then learn that R has occurred. In Situation 3 you initially believe that Q, and then learn that R has occurred.*

Postulate Statement: *The beliefs you have in Situation 1 are the same beliefs that would show up in either Situation 2 or Situation 3 (or both).*

Although participants did not give a reason for their choice, it is hypothesized that because Situation 1 includes both P and Q being true, and the inclusive disjunction of Situations 2 and 3 does not explicitly include the case where both P and Q are true, there may be a perceived difference in beliefs between the cases.

As a final note, the questions testing postulates U7 and U8 were the only two in which participants were asked to consider three situations, and as such included more text than the other postulates. As it has been documented that ease of information processing increases positive judgement [2], the increased cognitive load associated with these questions may have lead to a bias in participants towards disagreeing with them.

5.2 Section 2

Table 4 on the next page presents the results of Section 2 of the survey. For each postulate, the number and percentage of people who agreed or disagreed with the prescriptions of the postulate in question for that particular example is shown in the table. The final column of the table gives the most frequent reason for agreement with the postulates in question.

U5, U6, and U7 were tested using multiple questions. Referring back to Table 2 the reported rate of agreement for U6 corresponds to question 3 for the questions testing U6, for U5 it is question 3 for the questions testing U1, U5, and U8, and for U7 it is question 2 for the questions testing U7. It should be noted that the reported rates of agreement for all three of these questions were the lowest for that block of questions. For U1 the reported rate corresponds to question 2 for the block of questions testing U1, U5 and U8.

Taken as a whole it can be seen that participants broadly agreed with the prescriptions of the postulates for the examples in question, with the lowest rate of agreement with a postulate being 66.7%, that is U4. Given that U4 was perhaps the most cognitively demanding in that it involved both of De Morgan's Laws, this is perhaps not surprising.

Analysis of the reasons given for participants' agreement with the examples in question has revealed a surprising result. In all but the two cases (U5 and U7), the most frequent reason given by participants for their answers is directly related to the postulate in question. Or phrased another way, unless the participants had some non-formal variation of the postulate directly in mind, their answers do not make much sense. This is most obvious in the case of U4, U3 and U1, where the reasons given essentially directly match the postulates in question, but it also holds for the reasons given for U2 (nothing has changed, even although the new information is not equivalent to the old), U6 (both actions lead to the other) and U8 (per discussion in the Background section in this paper).

This is important for two reasons. First, even although only one instance of the given postulates is directly tested in these examples, the fact that the postulates or non-formal equivalents of them are referred to *as reasons* provides some evidence that the postulates hold more generally. This is as reasons are generally based on universal principles.

Second it seems to the author that there are two ways in which human reasoning could conform to the postulates in question. The first is as in U7 and U5 (which were not directly tested) here - the principles are followed, as in participants reason in conformance with them, but the way that they reason has nothing to do with the principles. The second way would be if the way people reason about these examples was in some way similar to the postulates. In the former case it might be said that while the postulates are a sufficient model from an external perspective, as an internal model of how people actually think they fall short; while in the latter this does not hold. The results of this section provide some evidence that this latter case does indeed hold for the majority of the postulates.

Table 3: Section 1 Responses

Postulate	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Combined Disagreement	Combined Agreement
U1	0 (0.0%)	3 (10.0%)	4 (13.3%)	**12 (40.0%)**	11 (36.7%)	3 (10.0%)	23 (76.7%)
U2	1 (3.3%)	6 (20.0%)	**8 (26.7%)**	**11 (36.7%)**	4 (13.3%)	7 (23.3%)	15 (50.0%)
U3	3 (10.0%)	7 (23.3%)	3 (10.0%)	**10 (33.3%)**	7 (23.3%)	10 (33.3%)	17 (56.7%)
U4	2 (6.7%)	2 (6.7%)	3 (10.0%)	**12 (40.0%)**	11 (36.7%)	4 (13.3%)	23 (76.7%)
U5	3 (10.0%)	11 (36.7%)	**1 (3.3%)**	**9 (30.0%)**	6 (20.0%)	14 (46.7%)	15 (50.0%)
U6	2 (6.7%)	3 (10.0%)	5 (16.7%)	**9 (30.0%)**	11 (36.7%)	5 (16.7%)	20 (66.7%)
U7	2 (6.7%)	**13 (43.3%)**	**4 (13.3%)**	5 (16.7%)	6 (20.0%)	15 (50.0%)	11 (36.7%)
U8	3 (10.0%)	9 (30.0%)	**5 (16.7%)**	10 (33.3%)	3 (10.0%)	12 (40.0%)	13 (43.3%)
Total	16 (6.7%)	54 (22.5%)	33 (13.8%)	**78 (32.5%)**	59 (24.6%)	70 (29.2%)	137 (57.1%)

Table 4: Section 2 Responses

Postulate	Agree	Disagree	Most frequent reason for agreement (number of people)
U1	27 (90.0%)	3 (10.0%)	New information should be believed (n=18)
U2	23 (76.7%)	7 (23.3%)	Nothing has changed (n=19)
U3	25 (83.3%)	5 (16.7%)	Transition from old information to new is possible, or give a possible scenario of such a transition (n=15)
U4	20 (66.7%)	10 (33.3%)	Both situations mean the same thing (n=18)
U5	26 (86.7%)	4 (13.3%)	New information should be believed (n=17)
U6	23 (76.7%)	7 (23.3%)	One light is on if the other is on, due to both lights sharing the same switch (n=18)
U7	28 (93.3%)	2 (6.7%)	New information should be believed (n = 22)
U8	26 (86.7%)	4 (13.3%)	Magazine may have been on the table before book was moved to the table (n=10)

5.3 Section 3

Table 5 on the next page shows the quantitative results of section 3. The first column of the table indicates which purported counter-example the question is part of, the next is a statement of propositional logic which indicates what the question was testing, and the final two columns show the respective number of people that agreed or disagreed with the statement, with the mode value indicated in bold.

From the table it can be seen that for each counter-example, for each statement in that counter-example, a majority of participants agreed with the responses as would be expected if they followed the form of the examples as outlined in Section 3 of this paper. Each counter-example, and the qualitative aspects of responses to them, are discussed in more detail below.

5.3.1 Counter-example 1.

Counter-example 1 asked the respondents to either agree or disagree whether $(p \land (q \lor \neg q)) \diamond ((p \land q) \lor (p \land \neg q) \lor (\neg p \land q)) \models p$. The atoms and new information were as interpreted as in the recurring robot example. Note that as $(p \land (q \lor \neg q))$ is equivalent to p and $((p \land q) \lor (p \land \neg q) \lor (\neg p \land q))$ is equivalent to $p \lor q$ this is a restating of the first objection to the KM postulates considered in Section 3 (U2 remains sufficient to derive the result). Overall 63.3% of participants disagreed with the statement, as would be expected given previous discussion in this paper.

The majority of disagree answers (n=16) when asked to give a reason for their choice gave some variation of the statement that $\neg q$ was in one of the possible states it were permissible for the robot to leave the office in. Within this three participants explicitly acknowledged that it was due to epistemic limitations about exactly how the robot carried out its task that they could not conclude that p. So generally the disagree answers agree with the theoretical discussion earlier in this paper.

Similarly the majority of agree answers (n=10) gave some variation of the statement that in the example it was stated that they initially remembered that p. This could be interpreted in two ways. First, as the information that you initially remembered p came early in the example, and the example text was rather long, respondents who disagreed may have missed that $\neg p \land q$ was specified as one of the possible tasks. Second, it could be interpreted as indicating that participants attributed *laziness* to the robot (as in Section 3), or that given that p is already the case, the robot would not change the room to ensure that $\neg p$. Three respondents explicitly mentioned that although $\neg p$ held in one of the states, the only uncertainty was regarding q, so for those three the second interpretation is almost certainly the correct one.

5.3.2 Counter-example 2.

Counter-example 2 was designed to test whether participants felt that $(\neg r \land \neg s) \diamond r$ was equivalent to $(\neg r \land \neg s) \diamond ((r \land s) \lor (r \land \neg s))$. The interpretation of the atoms was again as in the robot example. Overall 90% of participants agreed that the former models $\neg s$, while 83% of participants disagreed in the latter case. A majority of participants (n=25) gave as their reason for agreeing with the former statement that the robot was not instructed to change anything with regards to s.

When it came to reasons for disagreeing with the latter statement, a majority of participants (n=24) gave as their reason for disagreeing that one of the states it was permissible for the robot to leave the room in had it that $\neg s$. Because in the initial information it was specified that you remembered $\neg s$, for the reasons given by participants, this question should be analogous to Counter-Example 1. So the answers here suggest that for many participants the first

Table 5: Section 3 Responses

Counter-Example	Statement	Agree	Disagree
1	$(p \wedge (q \vee \neg q)) \diamond ((p \wedge q) \vee (p \wedge \neg q) \vee (\neg p \wedge q)) \models p$	11 (36.7%)	**19 (63.3%)**
2	$(\neg r \wedge \neg s) \diamond r \models \neg s$	**27 (90.0%)**	3 (10.0%)
2	$(\neg r \wedge \neg s) \diamond ((r \wedge s) \vee (r \wedge \neg s)) \models \neg s$	5 (16.7%)	**25 (83.3%)**
3	$(\neg c \wedge \neg d) \diamond c \models \neg d$	**26 (86.7%)**	4 (13.3%)
3	$(\neg c \wedge \neg d) \diamond d \models \neg c$	**26 (86.7%)**	4 (13.3%)
3	$(\neg c \wedge \neg d) \diamond (c \vee d) \models c \oplus d$	12 (40.0%)	**18 (60.0%)**
4	$(b \wedge \neg a) \diamond \neg b \models \neg a \wedge \neg b$	**24 (80.0%)**	6 (20.0%)
4	$(\neg b \wedge a) \diamond \neg b \models a \wedge \neg b$	**21 (70.0%)**	9 (30.0%)
4	$((b \oplus a) \diamond \neg b) \models \neg a \wedge \neg b$	**22 (73.3%)**	8 (26.7%)

interpretation of the agree answers to Counter-Example 1 is the correct one.

That $(\neg r \wedge \neg s) \diamond r$ is equivalent to $(\neg r \wedge \neg s) \diamond ((r \wedge s) \vee (r \wedge \neg s))$ follows either from U4, or from U6 and U1. So the counter-example is a counter-example to U4 and U6, or U4 and U1, or U4 and U6 and U1. However, given the reasons that participants gave for their answers (that the robot were not instructed anything regarding the atom s, and that one of the possible states of the room had it that $\neg s$), it does not seem remotely plausible that participants thought of this as a counter-example to U1. So it should be concluded that this is a counter-example to U4 and U6.

5.3.3 *Counter-example 3.*

This example was based on the intuition that even if $\phi \diamond \alpha \models \neg\gamma$ and $\phi \diamond \gamma \models \neg\alpha$, $\alpha \wedge \gamma$ could still be a model of $\phi \diamond (\alpha \vee \gamma)$. 86.7% of participants agreed with the preconditions for the question (i.e. the first two statements), with the majority of the reasons (n = 23, n = 24 respectively) being in both cases that for the example c holding had no impact on d or visa-versa. Similarly a majority of participants disagreed with the $c \wedge d$ is a model of the final statement. This was also the predominant reason (n = 20) for their disagreement with the statement (it should be noted that the way the question was phrased made it particularly obvious that $c \wedge d$ is typically considered a model $c \vee d$).

This example was the one with the lowest rate of agreement by participants. Reasons for not holding to the example generally seemed to be based on confusing whether $c \wedge d$ should be believed following the new information, as opposed to being believed as one possible state (the answers often talk of how although it could happen, they choose the other as it is not likely). It should be noted however, that the rate of agreement with this example was similar to that of the example testing postulate U4 in section 2, which involved a straight-forward application of De Morgan's Laws. It is thus likely that the rate of agreement with the example is within an acceptable error bound, given the error rate for the sample in the example testing postulate U4.

5.3.4 *Counter-example 4.*

Counter-example 4 was Lang's [22] objection to U8. Overall 80% of participants agreed that in the example $(b \wedge \neg a) \diamond \neg b \models \neg a \wedge \neg b$, 70% agreed that $((b \oplus a) \diamond \neg b) \models \neg a \wedge \neg b$, and 73% agreed that $((b \oplus a) \diamond \neg b) \models \neg a \wedge \neg b$. In terms of reasons given for agreement, 20 participants for the first statement, and 19 participants for the second statement, either stated the new information that b should have no impact on a, or gave the reason for their belief on the status of a as based on the initial beliefs, while the status of b was based on new beliefs, from which it could be inferred that they held b independent of a.

73% of participants agreed with the final statement that $((b \oplus a) \diamond \neg b) \models \neg a \wedge \neg b$, contra to the predictions of U8 given agreement to the previous two statements. The majority reason (n = 18) given for agreement with the statement was a variation of since you initially believed only one of Alice and Bob were in the office (modelled by $a \vee b$), and you saw one person leave, neither must be in the office now. So as Lang predicted, the new information in this case is seen as compatible with only one of the previously possible models of the world, mandating that U8 is inappropriate.

5.4 Section 4

Table 6 on the next page shows the results to all but two of the questions in Section 4. It is in the same format as Table 3. It can be seen from the table that in terms of the evaluation measures for the survey, participants generally agreed that the survey met these measures, with a median value of 'Agree' or higher for all the statements in question, and combined agreement that the survey met these measures being over 50% for all statements.

That the statement that the concrete questions were similar to examples that participants would come across in their everyday lives received the least overall support (63%) is presumably down to three of the examples using variations of the robot and the book example. However participants' strong rate of agreement with the statement that they reasoned in a manner similar how they would to everyday life indicates that they did not see this as a significantly impacting example on the applicability of their conclusions to their everyday manner of reasoning.

6 CONCLUSIONS

First the research questions are explicitly discussed. The first research question asked how intuitive people found the KM postulates. Based on the results of section 1 of the survey, the answer is that it depends on the postulate in question. U1, U3, U4 and U6 had a median value of agreed with by participants, so based on the results it can be concluded that participants found them intuitive. U2, U5 and U8 had a median value that did not seem significantly different from a neutral rating. Finally, U7 had a median value of

Table 6: Section 4 Responses

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Combined Disagreement	Combined Agreement
The instructions in the survey were clear and coherent.	0 (0.0%)	1 (3.3%)	1 (3.3%)	**18 (60.0%)**	10 (33.3%)	1 (3.3%)	28 (93.3%)
The concrete questions were similar to examples I would come across in everyday life.	0 (0.0%)	5 (16.7%)	6 (20.0%)	**9 (30.0%)**	10 (33.3%)	5 (16.7%)	19 (63.3%)
In the concrete questions I reasoned in a manner similar to how I reason in the everyday life.	0 (0.0%)	1 (3.3%)	5 (16.7%)	4 (13.3%)	**20 (66.7%)**	1 (3.3%)	24 (80.0%)
The abstract questions were understandable	3 (10.0%)	3 (10.0%)	3 (10.0%)	**12 (40.0%)**	9 (30.0%)	6 (20.0%)	21 (70.0%)

disagree, and thus it seems like it was the only postulate in which participants thought that it was actively *unintuitive*.

The results from section 2 of the survey indicate that the answer to the research question as to whether there exist instances in which people reason in accordance with each postulate is in the affirmative, with the lowest rate of agreement with the prescriptions of a postulate being 66.7%. Additionally the qualitative results indicate that for all postulates except U7 and U5, that the postulates (or non-formal variants of them) additionally seem to factor in to part of the *reason* why people seem to agree with the examples. They thus form evidence that for the postulates in question, the postulates may be indicative of a more general pattern of reasoning in human cognition.

The results from section 3 of the survey indicate that although there do exist instances in which people reason in accordance with the postulates, and there is some evidence that the postulates may form a more general pattern of reasoning, there also are instances in which the postulates are disobeyed in human reasoning. So the answer to the third research question is in the affirmative. This is most evident for U4 and U6 which were tested by counter-example 2. As about 70% of people also agreed with Langs' counter-example against U8, there is also evidence that U8 is violated. Finally, there is weaker evidence that U2 and one of U1, U4 and U5 are violated, as tested by counter-examples 1 and 5. Importantly, for all the counter-examples considered, the predominant reason people gave for their agreement or disagreement with the questions matched that as would be predicted by theory.

Second, the results from section 2 of the survey provide some evidence as to how to interpret section 1 of the survey. Note that generally (excepting U2), the postulates that received less qualitative support (U5, U7, U8) all received less support from participants as intuitive (n was only 10 in the case of U8 for qualitative support). The results therefore weakly suggest a reason as to why the postulates received less support as intuitive; that they are less available as reasons, or do not form an as entrenched pattern of reasoning. It should be noted that in the case of U7, as the postulate was not directly tested, this conclusion is not as compelling. However, given the form of U7 (reference to complete states of the world and conjunction of two separate belief update operations), it is intuitively reasonable that such a conclusion holds.

Finally two conclusions based on Section 3 of the survey are drawn. Generally the qualitative results from this section followed a pattern, in that *dependence* was seen as more important than minimizing change in beliefs (witness that the predominant reason given for e.g. participants' agreeal with $(\neg c \land \neg d) \diamond d \models \neg c$ is not that this leads to a minimal change in beliefs as it would in the PMA, but that c had no impact on d). This extends to the impact of new information on the knowledge base - e.g. in counter-example 2 the predominant reason given for $(\neg r \land \neg s) \diamond r$ modelling $\neg s$ was that s did not depend on r, while in $(\neg r \land \neg s) \diamond ((r \land s) \lor (r \land \neg s))$ not modelling s is that $(r \land s)$ was a model of the new information. The results of this section therefore provide some evidence that update operators that are based semantically on the notion of *dependence* (e.g. as in [11], [14]. [15]) as opposed to minimal change may be a better fit for modelling human reasoning than the KM approach.

The last conclusion to be drawn is that of how to interpret the results of the counter-example to U8. As noted in [16], U8 is considered what is essential to belief update vs. belief revision, such that if an operator does not satisfy U8, then it is not considered an update operator. So following Lang [22], it is agreed that while the example is a counter-example to U8 as it is commonly been presented and has been tested here, the import of the example should be a reconceptualization of where belief update is appropriate. Lang proposes that for any change that occurs in the world, we can consider separately both its *ontic* (as in effect on the world), and *epistemic* (as in its impact on our previously held beliefs) effects. He proposes that belief update is appropriate only in the case that the change in the world has only ontic effects on our beliefs, and no epistemic effects. As the majority reason given for participants' agreement with the last question of counter-example 4 in Section 3 is that the new information invalidates a previously thought possible model of the world, and that U8 runs counter to this recommendation, the results of this project provide some evidence that such a conception of belief update is appropriate for human reasoning.

7 LIMITATIONS AND FUTURE WORK

A recurring limitation of the work is that the survey questions may have been too complex given the length of the survey, such that the quality of responses may have been compromised by this. However the major limitation of this work is that it uses only a small sample of participants. So the generality of the results is in question, as it is not certain that these 30 participants would be representative of how people reason as a group. Future work might build on the results of this project by polling a larger group of people.

REFERENCES

[1] C. Alchourrón, P. Gärdenfors, and D. Makinson. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50 (1985), 510–530.

[2] A.L. Alter and D.M. Oppenheimer. 2009. Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* 13, 3 (2009), 219–235.

[3] M. Ben-Ari. 2012. *Mathematical Logic for Computer Science.* Springer Press, New York.

[4] A.J. Berinsky, G.A. Huber, and G.S. Lenz. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.comâĂŹs Mechanical Turk. *Political Analysis* 20 (2012), 351–368.

[5] A. Brawley. 2018. Using Amazon's Mechanical Turk for Psychology Research. http://abrawley.sites.gettysburg.edu/wp-content/uploads/2018/06/Using-Amazon's-Mechanical-Turk.pdf

[6] G. Brewka and J Halzberg. 1993. How to Do Things with Worlds: on Formalizing Actions and Plans. *Journal of Logic and Computation* 3, 5 (1993), 517–532.

[7] M. Buhrmester. 2018. M-Turk Guide. https://michaelbuhrmester.wordpress.com/mechanical-turk-guide/

[8] M. Buhrmester, T. Kwang, and S.D. Gosling. 2011. Amazon's Mechanical Turk:A New Source of Inexpensive,Yet High-Quality, Data? *Perspectives on Psychological Science* 6, 1 (2011), 3–5.

[9] M. Buhrmester, S. Talaifar, and S.D. Gosling. 2018. An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science* 13, 2 (2018), 149–154.

[10] M.J.C. Crump, J.V. McDonnell, and T.M. Gureckis. 2013. Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE* 8, 3 (2013), 1–18.

[11] P. Doherty, W. Lukaszewicz, and E. Madalińska-Bugaj. 1998. The PMA and relativizing change for action update. In *Proceedings of the International Conference on the Principles of Knowledge Representation and Reasoning*, A.G. Cohn, L. Schubert, and S.C. Shapiro (Eds.). Trento, 258–269.

[12] N. Friedman and J.Y. Halpern. 1999. Modeling Belief in Dynamic Systems Part II: Revision and Update. *Journal of Artificial Intelligence Research* 10, 1 (1999), 117–167.

[13] G. Gentzen. 1932. Über die existenz unabhängiger axiomensysteme zu unendlichen satzsystemen. *Math. Ann.* 107 (1932), 329–350.

[14] A. Herzig, J. Lang, and P. Marquis. 2013. Propositional update operators based on formula/literal dependence. *ACM Transactions on Computational Logic* 14, 3 (2013), 1–31.

[15] A. Herzig and O. Rifi. 1998. Update Operations: a Review. In *Proceedings of the 13th European Conference on Artificial Intelligence*, H. Prade (Ed.). John Wiley and Sons, 13–17.

[16] A. Herzig and O. Rifi. 1999. Propositional belief base update and minimal change. *Artificial Intelligence* 115, 1 (1999), 107–138.

[17] H. Katsuno and A. Mendelzon. 1991. Propositional Knowledge Base Revision and Minimal Change. *Artificial Intelligence* 3, 52 (1991), 263–294.

[18] H. Katsuno and A. Mendelzon. 1992. On the difference between updating a knowledge base and revising it. In *Belief revision*, P. Gärdenfors (Ed.). Cambridge University Press, 183–203.

[19] A.M Kellet and M. Winslett. 1985. On the use of an extended relational model to handle changing incomplete information. *IEEE Transactions on Software Engineering, SE-11* 7 (1985), 620–633.

[20] S. Kraus, D. Lehmann, and M. Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44 (1990), 167–207.

[21] Jon Krosnick and Stanley Presser. 2009. Question and Questionnaire Design. *Handbook of Survey Research* (03 2009).

[22] J. Lang. 2007. Belief update revisited. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI 07)*. IJCAI Press, 1534 –1540.

[23] Gabriele Paolacci and Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science* 25 (2014), 184–188.

[24] F.J. Pelletier and R. Elio. 2005. The Case for Psychologism in Default and Inheritance Reasoning. *Synthese* 146, 2 (2005), 7–35.

[25] M. Pittman and K. Sheehan. 2016. Amazons Mechanical Turk a Digital Sweatshop? Transparency and Accountability in Crowdsourced Online Research. *Journal of Media Ethics* 31, 4 (2016), 260–262.

[26] M. Ragni, C. Eichhorn, T. Bock, G. Kern-Isberner, and A. Ping Ping Tse. 2017. Formal Nonmonotonic Theories and Properties of Human Defeasible Reasoning. *Minds & Machines* 27 (2017), 79–117.

[27] M. Ragni, C. Eichhorn, and G. Kern-Isberner. 2016. Simulating human inferences in light of new information: A formal analysis. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI 16)*, S. Kambhampati (Ed.). IJCAI Press, 2604–2610.

[28] T. Shear and B. Fitelson. 2018. Two Approaches to Belief Revision. *Erkenntnis* (2018).

[29] G.M. Sullivan and Jr A.R. Artino. 2013. Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education* 5, 4 (2013), 541âĂŞ542.

[30] A. Tarski. 1956. *On some fundamental concepts of metamathematics. [1930] Logic, Semantics, Metamathematics. Papers from 1923 to 1938, translated by J.H. Woodger. Pages 30–36.* Clarendon Press.

[31] M. Winslett. 1988. Reasoning about action using a possible models approach. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI 88)*. AAI Press, 89–93.