0.1 Section 3

Table 1: Counter-Examples

Counter-Example	Statement
1	$(p \land (q \lor \neg q)) \diamond ((p \land q) \lor (p \land \neg q) \lor (\neg p \land q)) \models p$
2	$(\neg r \land \neg s) \diamond r \models \neg s$
2	$(\neg r \land \neg s) \diamond ((r \land s) \lor (r \land \neg s)) \models \neg s$
3	$(\neg c \land \neg d) \diamond c \models \neg d$
3	$(\neg c \land \neg d) \diamond d \models \neg c$
3	$(\neg c \land \neg d) \diamond (c \lor d) \models c \oplus d$
4	$(b \land \neg a) \diamond \neg b \models \neg a \land \neg b$
4	$(\neg b \land a) \diamond \neg b \models a \land \neg b$
4	$((b \oplus a) \diamond \neg b) \models \neg a \land \neg b$

Table 1 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.

0.1.1 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 $\neg p \wedge q$ 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.

0.1.2 Counter-example 2

Counter-example 2 was designed to test whether participants felt that $(\neg r \wedge \neg s) \diamond r$ was equivalent to $(\neg r \wedge \neg s) \diamond ((r \wedge s) \vee (r \wedge \neg s))$. 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 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.

0.1.3 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.

0.1.4 Counter-example 4

Counter-example 4 was Lang's [LangLang2007] 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.

References

[LangLang2007] J. Lang. 2007. Belief update revisited. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI 07)*. IJCAI Press, 1534 –1540.