The Quality of Test Context and Contra-evidence as a Moderating Factor in the Belief Revision Process

Kristien Dieussaert (kristien.dieussaert@psy.kuleuven.ac.be)
University of Leuven, Department of Psychology
Tiensestraat 102, Leuven, Belgium

Walter Schaeken (walter.schaeken@psy.kuleuven.ac.be)
University of Leuven, Department of Psychology
Tiensestraat 102, Leuven, Belgium

Gery d'ydewalle (gery.dydewalle@psy.kuleuven.ac.be)
University of Leuven, Department of Psychology
Tiensestraat 102, Leuven, Belgium

Abstract

In this study, we describe the influences of qualitative changes to the reasoning problem on the reasoning process. The first manipulation is the quality of the test context: A rule is learned in a certain context and contradicted in another. The belief in the rule is then measured in the learning context, the contradictory context, and a new context. The second manipulation is the quality of contradiction: The contradictory rule can neutralize or inverse the learned rule. Both qualitative changes influence the belief revision process.

Introduction

Research in Artificial Intelligence is often conducted to develop systems that think/act rationally (or like humans, depending on the approach). Minsky (1975) was one of the first to point out the problem with deductive systems, and from then on, several researchers developed non-monotonic reasoning systems (for an overview, see Brewka, Dix, & Konolidge, 1997).

While these systems can be very interesting from the viewpoint of an engineer, Rips (1994) mentions two main reasons why these nonmonotonic logics are less than ideal for cognitive purposes: They do not lend themselves to simple implementations (higher order logics are incomplete) and they do not seem to reflect the deliberations that actually underlie human reasoning with defaults. He (1994, p. 299) stated: "For purposes of philosophy (and perhaps AI), we need normative rules about how to change our minds in the face of conflicting evidence; for purposes of psychology, we also need descriptive information on how people deal with the same sort of conflict."

Elio and Pelletier (1997) wrote a pioneering article on this topic. In their experiments, they first presented participants with a conditional premise and a categorical premise that affirmed the antecedent [or denied the consequent]. Then, they added a third piece of information, which conflicted with the conditional and the categorical premise. This piece of information was a categorical premise denying the consequent [or affirming the antecedent]. The three pieces of information together are in contradiction with the valid Modus Ponens [Modus Tollens] argument: If A then B, A, thus B [If A then B, not-B thus not-A].

Participants were asked to resolve the contradiction by rejecting one of the two first pieces of information. Elio and Pelletier (1997) observed that participants chose to disbelieve the conditional premise rather than the categorical one when resolving the contradiction.

Their results were refined by Dieussaert, Schaeken, De Neys, and d'Ydewalle (2000) and by Politzer and Carles (2001). They found that the initial belief in the conditional premise influenced the belief revision choice that participants prefer to make. When participants had a strong belief in the conditional premise, the preferred to reject (doubt) the categorical premise, when the conflicting information is added. When participants had a weak belief in the conditional premise, the results of Elio and Pelletier (1997) were confirmed: In this case, participants prefer to reject (doubt) the conditional premise.

This shows that it is important in belief revision research to be aware of the belief state participants hold before conflicting information is presented. Therefore, we conducted a measure of the initial belief state in the following experiments, before adding conflicting information.

The following experiments were inspired by research in the field of conditioning. Among others, Bouton (e.g., 1988, 1994) showed that the extinction of behavior does not necessarily means the rejection of a learned rule. One of the phenomena that confirm this hypothesis is 'renewal': When a behavior is learned in context A, and extinguished in context B, the behavior might show up again with a new test in context A. Bouton explains this phenomenon as follows: Individuals learn a dominant rule and exceptions to this rule in certain contexts (see also: Holyoak, Koh, &

Nisbett, 1989). As a consequence, a stimulus that has lost his value as a reinforcer, becomes an ambiguous stimulus from which the specific value is determined by the context. In other words, Bouton points to the importance of the presence of a certain context as an indicator of the belief in the (conditional) rule when an individual is confronted with conflicting information to that rule. In the forthcoming experiments, we will also work with context embedded situations, to gain more insight in the relative value of the contra-evidence.

Another uniqueness of the present experiments is that the participants' belief state is measured by the behavior they pose, and not by what they say their belief is. We think it is as important to know how individuals act on their beliefs, as it is to know how they describe their beliefs. Moreover, De Neys, Dieussaert, Schaeken, and d'Ydewalle (2000) found a strong correspondence between what participants say about their beliefs and how they actually act upon these beliefs, in experiments comparable with the ones presented below.

Briefly, on a technical level, the following experiments are designed as follows: first, the initial belief in a rule is tested. Then, contradictory evidence is presented in a different context, and finally, the belief in the rule is again evaluated by examining participants' manifested behaviours. On the content level, the experiments describe the influence of qualitative changes to the reasoning problem on the belief revision process.

Experiment 1: Pilot study

In this part, we describe a study that tests whether formerly used instantiations of the quality of contra-evidence were well chosen.

Before that, we explain the two hypotheses that are at issue in this manuscript and we depict the important role that the pilot study has regarding the second hypothesis. Subsequently, we give a brief overview of the content of the experiments.

Two hypotheses are evaluated in this manuscript. The main hypothesis is that when a conditional rule is acquired in a first context (A) and contradicted in a second context (B), the belief in the conditional is not affected by the contradictory information when tested in the first context (A) or in a new context (C). Indeed, if the rule acquired in the first context really is a dominant rule, and if the conflicting information learned in a second context is perceived as exceptional, one should also apply the dominant rule in a new context.

A second hypothesis is that the quality of the contradictory information plays a role in the belief revision process. Bouton (1988) focused on the *extinction* of learned rules. With extinction, a learned rule is often only neutralized (e.g. food – no food). Our hypothesis is that when the conflicting information has a stronger impact (compare with: food – shock), this may affect the belief revision process in a different way. More precisely, we hypothesize that when the contradictory information merely neutralizes the acquired conditional rule, its effect on the belief revision process is smaller than when it reverses the acquired conditional rule. Or, in other words, we suppose that when the acquired rule is reversed, this might affect the final belief state tested in context A and C anyway, despite the fact of a buffering context element (B) when the contradictory information is provided.

It is made very clear in the fore mentioned examples that the absence of an appetitive stimulus (neutralizing) is not the same as the presence of an aversive stimulus (reversing). An adequate test of the second hypothesis is only possible if both stimuli are well chosen.

In a former series of experiments (Dieussaert, Schaeken, & d'Ydewalle, 2002), the manipulation of the quality of the contra-evidence did not result in significant differences, contrary to our second hypothesis. Before drawing any theoretical consequences from this finding, we examine whether those instantiations were appropriately chosen. The pilot study (Experiment 1) was set up to test this hypothesis.

To convey a good understanding of this pilot study and its consequences possible, we briefly describe the content of the former experiments (Dieussaert et al., 2002). The motivation behind the experiment is related to that behind the conditioning experiments of Pineño, Ortega, and Matute (2000). Participants are in a war area, leading a rescue mission. They are told they should rescue as many refugees as possible from a building and they can do this by loading the refugees on a truck. Importantly, they should only fill the truck with refugees if the road is free of mines. They can learn whether the road is safe because coloured lights indicate it. Therefore, they should learn as fast as possible the meaning of these lights.

The participants learn the meaning of the lights in a first location (context), and then move to a second location where they learn that the meaning of some lights is reversed or neutralized. For example, a green light in the first location might indicate that the road is safe, while in the second location it indicates that the road is mined (reversal) or that no information about the road is available (neutralizing). Finally, they move to a third location. At this location, identical or different from one of the former contexts, the participants' belief regarding the meaning of the lights is examined.

In this pilot study (Experiment 1), we manipulate the feedback that participants received in the test phase. Three forms of feedback were distinguished: clearly

positive feedback, clearly negative feedback, and the feedback 'no information available'. The hypothesis is that participants interpret the latter feedback as negative.

Method

Participants. The 36 participants are candidate students at the University of Leuven, Department of Psychology and they participated as a partial fulfillment of a course requirement. Each student was randomly attributed to one of two groups (12 per group).

Design One variable was manipulated between subjects: Feedback. The three levels of this variable were manipulated in the test phase.

The dependent variable is the number of times participants press on the space bar during a fixed time interval when the cue light appears. The meaning of a space bar press is that a person is put in the truck. Each press is thus equivalent to the saving of one person. This dependent variable is measured in the three phases of the experiment: the confirmation phase (9 blocks), the contradiction phase (9 blocks) and the test phase (4 blocks). This behavioral measure is taken as a measure of belief state. The idea behind it is that the more the participant believes that the road is free of mines, the more persons (s)he will put in the truck.

Each block consists of four trials. A trial lasts four seconds. During these four seconds, a colored light is shown (green, blue, red or white). The sequence of the lights is randomized within each block. Participants learn the meaning of these lights (see below) by pressing the spacebar. One of these lights is the cue light, one is the neutral light and the other two do not have a fixed meaning. Each light's meaning is counterbalanced between participants.

The first eight blocks are considered as learning blocks; they are not included in the analysis. All participants acquired a constant pressing level for the cue light within these blocks. The cue light indicates a safe road. The neutral light indicates that the road is closed. The meaning of the other lights differs randomly over blocks; they can indicate that the road is mined, safe, or closed. The context is a location, which we refer to as context A.

The confirmation phase consists of nine blocks. The meaning of the lights is equal as in the learning blocks. The context does not change either.

The contradiction phase consists of nine blocks. Here, the cue light indicates a mined road. The neutral light indicates that the road is closed. The meaning of the other lights differs randomly over blocks; they can indicate that the road is mined, safe or closed. The context is a location, which we refer to as context B.

The test phase consists of four blocks. The test phase always takes place in the same context as the confirmation phase (A), but the feedback varies between groups. For the ABA/+/ group, the message 'n persons saved' is displayed after the cue light lit up in the test phase, while for the ABA/-/ group and the ABA group, the messages are 'n persons died' and 'no information available', respectively. The neutral light indicates that the road is closed. The meaning of the other lights differs randomly over blocks; they can indicate that the road is mined, safe, or closed.

Procedure The experiment was carried out individually, on computer. All participants received the same instructions. They were told to imagine that they were in charge of a rescue mission. Refugees were hiding in a building that could be attacked by the enemy. Furthermore, these people can only be rescued safely, by placing them in a truck, if the road is free of mines. People are loaded in the truck by pressing the spacebar. To know whether the road is safe, the participants have to learn the meaning of the lights that appear on each trial. Some of the four lights have a fixed meaning (road safe, road mined, road closed); other lights are distracters and have no fixed meaning. The feedback they receive is 'n persons saved', 'n persons died', or 'road closed'.

Moreover, sometimes no information is available about the situation of the road, so that the participants do not know whether the persons they placed in the truck were saved. It was stressed that they should only press the spacebar continuously in case they were very sure that the people would be saved in that trial.

Finally, they were told that different rescue missions could take place and that a beep sound signaled the start of a new mission. The importance of keeping in mind the location (context) where the rescue mission took place was stressed. They were told that the lights could have different meanings in different locations.

Results

Table 1 shows the mean number of space bar presses, (i.e., persons put in the truck) within a fixed time interval. Neither in the confirmation phase, nor in the contradiction phase, did we observe any differences between the groups.

Table 1: Mean number of presses per phase per group.

	CONFIRMATION PHASE	CONTRADICTION PHASE	TEST PHASE
ABA/-/	34.9	7.1	14.9
ABA	45.5	2.5	27.5
ABA/+/	42.6	4.9	39.1

An ANOVA shows a main effect of Feedback ($F(2,33) = 7.80$, $p<.005$). The ABA group scores higher in the test phase than the ABA/-/ group ($F(1,33) = 4.20$, $p<.05$), but does not differ from the ABA/+/ group. The latter group differs strongly from the ABA/-/ group ($F(1,33) = 15.58$, $p<.0005$).

A more detailed planned comparison analysis on the four blocks of the test phase moderates the picture. The three groups do not differ from each other in the first block. The ABA group differs significantly from the ABA/-/ group in the second block (36.7 vs.9.4; $F(1,33) = 14.24$), $p<.001$), but differs significantly from the ABA/+/ group in the third and fourth block of the test phase (17.8 vs. 42; $F(1,33) = 10.54$, $p<.005$ and 18.7 vs.36.3; $F(1,33) = 5.51$, $p<.05$, respectively). For these last two blocks, no difference with the ABA/-/ group could be observed.

Discussion

The feedback message 'no information available' does not function well as a neutralizer of a conditional rule. The message is not interpreted as the absence of an appetitive stimulus, but rather as the presence of an aversive stimulus. Indeed, after a few trials the effect of this feedback message does not differ from other negative feedback, such as the message 'n persons died'.

Experiment 2

Now that we have shown that the message 'no information available' is interpreted as negative rather than as neutral feedback, we have good reasons to believe that it was not appropriate to use this message as an instantiation of the neutralizing level as was done in former experiments (see: Dieussaert et al., 2002).

In the present experiment (Experiment 2), we test the two hypotheses at issue, and focus on the more appropriate qualitative manipulations of the contra-evidence. Our main hypothesis is that a conditional rule acquired in a first context functions as a dominant rule. Conflicting information learned in a second context is perceived as exceptional. Therefore, one should also apply the dominant rule in a new context.

Our second hypothesis is that the quality of the contradictory information plays a role in the belief revision process. More precisely, we suppose that when the acquired rule is reversed by contradictory evidence, this might nonetheless affect the final belief state in both context A and context C, despite the existence of the buffering context element (B).

Method

Participants Seventy candidate students at the University of Leuven, Department of Psychology, participated in this experiment. They did not participate in the former experiment and participated as a partial fulfillment of a course requirement. Each student was randomly assigned to one of six groups (11 in the ABB- and ABC-Weak group; 12 in the other groups).

Design The independent variables, Test Context and Contradiction Level, were manipulated between subjects. We distinguish three instances of the variable Test Context: The context in the test phase can be the same as the context in the confirmation phase (ABA), it can be the same as the context in the contradiction phase (ABB) or it can differ from both (ABC).

The second independent variable, Contradiction Level, consists of two instances: The contradiction can be Strong, when the meaning of the rule is reversed, or can be Weak, when the meaning of the rule is neutralized.

The dependent variable is the same as in Experiment 1: The number of times participants press on the space bar during a fixed time interval when the cue light appears. Again, this variable is measured in the three phases of the experiment: the confirmation phase (9 blocks), the contradiction phase (9 blocks) and the test phase (4 blocks). This behavioral measure is taken as a measure of belief state.

Each block consists of four trials. A trial lasts four seconds. During these four seconds, a colored light is shown (green, blue, red, or white). The sequence of the lights is randomized within each block. Participants learn the meaning of these lights (see below) by pressing the spacebar. One of these lights is the cue light, one is the neutral light and the other two do not have a fixed meaning. Each lights' meaning is counterbalanced over the participants.

The first eight blocks are considered as learning blocks; they are not included in the analysis. All participants acquired a constant pressing level for the cue light within these blocks. The cue light indicates a safe road. The neutral light indicates that the road is closed. The meaning of the other lights differs randomly over blocks; they can indicate that the road is mined, save, or closed. The context is a location, which we refer to as context A.

The confirmation phase consists of nine blocks. The meaning of the lights is equal as in the learning blocks. The context does not change either.

The contradiction phase consists of nine blocks. Here, the cue light indicates a mined road in the Strong contradiction level. In the Weak contradiction level, the message 'road closed' is displayed with a cue light. The neutral light indicates that the road is closed. The meaning of the other lights differs randomly over blocks; they can indicate that the road is mined, save or closed. The context is a location, which we refer to as context B.

The test phase consists of four blocks. In this phase, simultaneously with the cue light, the message 'no information is available' is displayed. The neutral light indicates that the road is closed. The meaning of the other lights differs randomly over blocks; they can indicate that the road is mined, save, or closed. The context varies over groups: It can be the same as in the confirmation phase (A), the same as in the contradiction phase (B) or different from those two (C).

Procedure See Pilot study (Experiment 1)

Results

No differences between the groups could be observed in the confirmation phase. This was expected since at that moment no manipulation was introduced yet. In the contradiction phase, the Strong groups scored lower than the Weak groups, as we hypothesized (2.4 vs.7.6; $F(1,64) = 7.6$, $p<.01$). Within the same level of contradiction, no differences were observed. Table 2 shows the means for each group.

Table 2: Mean number of presses per phase per group.

		CONFIR-MATION PHASE	CONTRA DICTION PHASE	TEST PHASE
ABA	Strong	45.5	2.5	27.5
ABB	Strong	41.2	1.8	2.1
ABC	Strong	41.0	2.8	2.5
ABA	Weak	44.0	10.9	40.4
ABB	Weak	44.0	8.3	7.3
ABC	Weak	46.2	3.6	2.9

An ANOVA shows a main effect of Test Context ($F(2,64)=73.21$; $p<.00001$). The score of participants from the ABA group (33.9) is higher than the score of participants from the ABC group (2.69; $F(1,64) = 118.36$, $p<.00001$) and the ABB group (4.70; $F(1,64) = 98.65$, $p<.00001$).

The manipulation of Contradiction Level resulted in a main difference between both levels (10.69 vs. 16.87 for the Strong and Weak contradictions respectively; $F(1,64) = 6.71$, $p<.05$) in the expected direction. Neither the ABB group, nor the ABC group has a different score in the Strong and Weak group. The main effect results from the difference in the ABA Strong and Weak group ($F(1,64) = 10.05$, $p<.005$).

The ABA Weak score in the test phase does not differ from it's score in the confirmation phase, while the ABA Strong score does ($F(1,64) = 19.79$, $p<.00005$).

Discussion

The effect of Contradiction Level is now adequately obtained due to an appropriate choice of the instantiation of the Weak level. Only the groups where learning and test context are the same, are affected by the Contradiction Level manipulation: When the contradiction merely neutralizes (i.e., does not reverse) a previously learned rule, no belief revision takes place. When the contradiction is strong, the belief in the rule decreases compared to the learning phase. This observation is surprising given that the contradiction takes place in another context, and could therefore be easily neglected.

On a theoretical level, it would be tempting to conclude that the rule acquired in context A can be considered as a dominant rule. One could state that the first and the second hypothesis are thus confirmed: The belief is not influenced by the contradictory information in context B, when this conflicting information neutralizes the learned information, but it is influenced when the conflicting information reverses the learned information.

However, this conclusion would be false because for each Contradiction level, the belief in the rule remains low when it is tested in a new context (C). If the 'dominant rule' hypothesis were correct, the belief in the new context should be as high as in the confirmation phase, at least for the Weak level group.

General Discussion

We manipulated two qualitative factors, Test Context and Contradiction Level, to test their influence on the belief revision process. Regarding Test Context, we hypothesize that when a conditional rule is acquired in a first context (A) and contradicted in a second context (B), the belief in the conditional is not affected by the contradictory information when subsequently tested in the first context (A) or in a new context (C).

This hypothesis was inspired by Bouton (1988, 1994), who stated that an acquired rule is protected against extinction when the conflicting information is presented in another context than the learning context, because individuals interpret the first rule as the dominant rule and other one as the exception.

Regarding Contradiction level, the hypothesis was that the quality of the contradictory information plays a role in the belief revision process. We supposed that when the acquired rule is reversed, this might affect the final belief state tested in context A and C anyway, despite the buffering context (B) of the contradictory information.

The influence of the variable Test Context is clear: Given the same learning (and confirmation) and test context, less belief revision takes place after contradictory information is presented in a different context, than when learning and test contexts differ.

The influence of the variable Contradiction Level is also clear. Given an appropriate choice of the absence of an appetitive stimulus (road closed) and of the presence of an aversive stimulus (road mined), the manipulation resulted in a significant difference between the Weak and the Strong level. This effect was mainly due to the effect in the ABA group. For this group, no belief revision took place when the contradiction neutralized the learned rule, but belief revision was observed when the contradiction inversed the learned rule.

Although this effect on the ABA group confirms both hypotheses, one should not be too optimistic about the theoretical consequences because no such effect was

found for the ABC group, which plays an important role as a control group.

Therefore, the results may put serious question marks on the theoretical translation into 'dominant' and 'exception' rules. The results of these experiments may indicate that different rules are learned within each context.

However, from an economical viewpoint on the learning process, this idea does not seem very fruitful. Learning and applying conditional rules would become a very heavy task if every new context implied a new rule.

An interesting distinction, that was not made thus far, could be between conditional rules that are learned (and mostly applied) in one restricted context and conditional rules that are learned (and applied) in a wide range of different contexts.

Our experiments investigated the former situation. It might be that in the latter situation rules that are applicable in more contexts may become dominant rules and rules that are only applicable in one or a few contexts may become exceptional rules. Surely, this idea needs supplementary testing.

In sum, the results of these experiments indicate that the belief revision process is influenced by various qualitative characteristics of the problems. More precisely, this study shows the effect of contradictory context and the level of contradiction. Although the theoretical consequences of these results are not clear for the time being, some handouts for further testing are provided.

Sharing the considerations of Rips (1994) regarding human reasoning, we conclude with an amendment to his statement: For purposes of psychology, we need more descriptive information *and theoretical explanations* on how people deal with conflicting information.

Acknowledgments

This study was carried out with support from the IUAP/PAI P4 and from the Fund for Scientific Research Flanders Project G.0239.02 .

We wish to thank the reviewers and Uri Hasson for their helpful comments.

References

Bouton, M. E. (1988). Context and ambiguity in the extinction of emotional learning: implication for exposure therapy. *Behavior Research and Therapy, 26*, 137-149.

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior Processes, 20*, 219-231.

Brewka, G., Dix, J., & Konolidge, K. (1997). *Nonmonotonic reasoning: an overview*. Stanford: CSLI Lecture Notes 73.

De Neys, W., Dieussaert, K., Schaeken, W. & d'Ydewalle, G. (2000). *Assessing belief revision in a game context*. Poster presentation, 4th International Conference on Thinking, Durham.

Dieussaert, K., Schaeken, W., De Neys, W., & d'Ydewalle, G. (2000). Initial belief state as predictor of belief revision. *Current Psychology of Cognition*, 19, 277-288.

Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). *A study of the belief revision process: The value of context and contra-evidence* (Tech. Rep. N°291). Leuven, Belgium: University of Leuven, Laboratory of Experimental Psychology.

Elio, R., & Pelletier, F. (1997). Belief change as propositional update. *Cognitive Science, 21 (4)* , 419-460.

Holyoak, K. J., Koh, K., & Nisbett, R.E. (1989). A theory of conditioning: Inductive learning within rule-based default hierarchies. *Psychological Review, 96*, 315-340.

Minsky, M. L. (1975). A framework for representing knowledge. In Winston, P.H. (ed.), *The psychology of computer vision* (pp. 211-277). NY: McGraw-Hill.

Politzer, G., & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking and Reasoning, 7(3)*, 217-234.

Pineño, O., Ortega, N., and Matute, H. (2000). The relative activation of associations modulates interference between elementally trained cues. *Learning and Motivation, 31 (2),* 128-152.

Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: Routledge.