COUNTERFACTUAL REASONING:
RESOLVING INCONSISTENCY BEFORE YOUR EYES

Russell REVLIN, Dustin P. CALVILLO, and Stephanie BALLARD
University of California, Santa Barbara, USA

Belief revision has invariably been studied with abstract relationships using paper-and-pencil tasks. The present study asks whether people employ the same procedures in revising their beliefs when they concern concrete situations rather than arbitrary, abstract ones. Students listened to a narrative and then verified that selected statements from it were true and consistent. They then needed to revise their beliefs as a result of confronting a belief-contravening assumption that they had to accept as true even though it was false of the narrative. There are two main findings. First, when the artificial beliefs were part of an integrated, narrative structure, participants treated them as they do natural semantic categories and not as arbitrary relations. Second, when reasoning about concrete situations in front of them, reasoners readily changed the properties of the objects in order to retain the belief structure of the narrative. This suggests that college students act as if they were Platonists who view the observable world as fallible reflections of the idealized world.

When we revise our beliefs in the face of an assumption that we disbelieve, we are doing *counterfactual reasoning* - we are reasoning from false assumptions. How we revise our beliefs in these circumstances depends on the content of the material (see also Dieussaert, De Neys, & Schaeken, 2005; Ford, 2005). When the beliefs are artificial (e.g., *If an ancient ruin has a force field...* cf., Byrne & Walsh, 2002; Elio & Pelletier, 1997) and there is a conflict among beliefs, reasoners show only a modest preference for treating some beliefs as more important than others. In contrast, when beliefs are consistent with the reasoner's semantic knowledge (e.g., *All lions are mammals*), reasoners tend to resolve the inconsistencies by giving priority to long-held generalizations (e.g., Revlin, Cate, & Rouss, 2001; for an alternative view see Legrenzi & Johnson-Laird, 2005). This pattern of findings is based on paper-and-pencil tasks where the only consequence of belief-contravening assumptions is limited to the acceptance or rejection of statements in a list and has little impact on the reasoner's everyday world. The present study examines two aspects of belief revision: (a) under what conditions is it

sensible to investigate belief revision processes with artificial materials, which have no *apriori* believability; (b) do we revise our beliefs differently when their referents are only imagined rather than right in front of us?

Although people reason differently depending on whether or not the facts are artificial and/or arbitrary (e.g., Nguyen & Revlin, 1993; Revlis & Hayes, 1972; Wilkins, 1928), there is no independent standard of *artificiality* and how it relates to *believability*. Unless we have such a standard in place, we will not be sure when and how someone will decide which of their beliefs are candidates for revision. The present study explores the possibility that the critical factor is the degree to which the "facts" are integrated within a larger network of relations. When they are integrated, reasoners may treat those relations as entrenched as they do semantic categories (e.g., Smith, Haviland, Buckley, & Sack, 1972). But when the set of facts are not integrated, reasoners will treat them as a mere list of arbitrary relations with no particular believability or commitment. We simulate this integrated structure in Experiment 1 by presenting stories to the reasoners to determine whether non-believed material can temporarily evoke an investment of belief.

The often expressed Italian proverb, "seeing is believing" (*Chi non l'occhio vede, col cuor crede),* suggests that people will accord a high degree of commitment (i.e., *belief*) to relations they observe in the physical world. If this were correct, then the pattern of belief revision would reflect a greater weight given to statements of things the reasoner observes than to more abstract propositions not directly linked to what the reasoner sees - such as relations in a story. Alternatively, if reasoners adopt a more Platonic stance, they will see the concrete world as a fallible reflection of the abstract one and will revise inconsistencies in their beliefs in a way that preserves their abstract belief at the cost of what they actually see. To study how reasoners reconcile inconsistencies among propositions associated with physical objects, Experiment 2 contrasts the standard paper-and-pencil paradigm of Experiment 1 with one where there are physical consequences of revising one's beliefs.

The Paradigm

There are occasions when we revise our beliefs in light of a "what if" conjecture that introduces an inconsistency into those beliefs. This is the hallmark of counterfactual reasoning - reasoning from false assumptions. To investigate the cognitive processes involved in this form of reasoning, we employ a paradigm called *belief contravening problems* (e.g., Rescher, 1964). To illustrate, let us suppose that after reading a story about a mythical kingdom, the participant certifies his/her belief that the first three statements (1a) - (1c) are true of the story.

(1) (a) All of King William's knights have swords.
    (b) This knight has a battle-ax (not a sword).
    (c) This knight works for King Surly.
    (d) Assume: This knight works for King William.

These statements constitute the corpus of belief into which we introduce (1d) an assumption that is in clear violation of one of the beliefs (1c). To reconcile this inconsistency requires rejection of (1c). There is another inconsistency introduced by this assumption. If it is joined with (1a), the two together violate (1b). Alternatively, if the assumption is joined with (1b), the two together violate (1a). Which of these two possibilities will the reasoner prefer in order to have a consistent set of statements that includes the assumption?

Before we consider the answer to this, there is a second type of belief-contravening situation that also requires belief revision. It is shown in (2).

(2) (a) All of King William's knights have swords.
    (b) This knight works for King William.
    (c) This knight has a sword.
    (d) Assume: This knight has a battle-ax (instead of a sword).

Again, the assumption directly contradicts one of the statements (2c) (participants are aware that the knights cannot have both sword and battle-ax) and sets up two possible inconsistencies among the remaining set, just as in (1) above. The second problem differs from the first because the assumption in (1d) includes a new token (the knight) within the super-ordinate category *Employees of King William*. In contrast, the assumption in (2d) excludes an existing token (the knight) from having a necessary property of the superordinate (*swords*). The situation described in (1) we call a *combining problem* because it adds something to the relationship between the particular instance (1d) and the general statement (1a) and we call the situation (2) *rending* because it takes something away from that relationship. This nomenclature focuses on the role that the belief-contravening assumption plays in the problem. Labels used by other researchers for these types of problems focus on the logical structure of the pre-assumption set of statements and refer to them as Modus Tollens and Modus Ponens, for problems (1) and (2), respectively (e.g., Elio & Pelletier, 1997).

Briefly, there are three findings from prior research pertinent to the present study (summarized in Revlin, Cate, & Rouss, 2001). First, in combining problems such as (1), reasoners tend to revise their beliefs in the situation by joining the assumption with the generality (1a), causing them to reject the particular statement (1b). It is a bit surprising that there is a reliable prefer-

ence since there is no logically compelling reason to select one alternative over another. Second on rending problems such as in (2), reasoners are equivocal as to which choice to make. Finally, the *apriori* belief in the statements plays a role in the degree to which reasoners prefer one configuration over another. When problems like (1) and (2) contain artificial categories, as in the present examples, the tendency to choose one path over another is muted.

This last result motivates Experiment 1 to investigate the possibility that belief revision with artificial domains is different from belief revision with real-world relations because the latter is integrated and the former tends not to be. To examine this, students read a narrative that contains artificial relations, but relations that are integrated just as might be found with real-world categories. If the integration of categories is a sufficient condition for investment of belief, then the pattern of reasoning should be equivalent to what has been previously found with real-world content.

This experiment has a second purpose of providing a frame of reference for Experiment 2, where belief revision will be framed within a concrete situation. That experiment enables us to see whether the inferences drawn with concrete materials are the same as drawn with text-based reasoning.

Experiment 1

Method

Participants. The participants were 84 undergraduates from an introductory psychology class fulfilling a course requirement. They were run in sessions lasting about 15 minutes in groups of up to 5 students.

Materials. Two narratives were created as a joint effort by undergraduates. The first story describes two kingdoms with knights doing battle. The second describes a confrontation in a Wild West town between raiders and settlers. An example story is shown in the Appendix. The total set of problems was constructed from the orthogonal contrast of Assumption (Combining or Rending) and Relation (Class inclusion or Property-Assignment). With respect to Relation, half of the problems contained generalities that expressed class-inclusion relation (e.g., *All of King William's knights are residents of Fantasia*) and half expressed property-assignment (e.g., *All of King William's knights have strong swords*). Previous research (e.g., Revlis & Hayes, 1972) found that reasoners treat property relations as more mutable and less believable than class relations. We sought to replicate that finding here. Half of the generalities were affirmatively expressed (*All X are/have Y*) and the remaining half was negatively expressed (*No X are/have Y*).

Procedures. Participants read two, 300-word narratives. After reading the first story, the students solved a belief-contravening problem like those

shown above. They then read a second narrative and solved an accompanying problem. Prior to reading the first narrative, the participants learned about the nature of the reasoning task. For half of the students, the first problem contained an affirmative generality and the second problem contained a negative generality. The reverse was true for the remaining half of the students.

Results

In each problem, the reasoner's task was to decide which statements to retain and which to reject in order to restore consistency to the set of statements. Table 1 shows the percentage of trials on which the participants chose to accept the general statement and reject the particular statement. Overall, our students reject the particular statement on 68% of the trials, which is marginally greater than chance ($p < .05$). However, this does not reflect the quite different preferences that result from the two types of Assumptions. Students reliably prefer to retain the general statement more often on Combining problems than on Rending ones [$F(1, 80) = 5.4$, $p<.05$], which replicates previous studies mentioned above. As revealed in Table 1, no other main effects or interactions were shown in this study.

Table 1.
Percent Preference for Retaining the General Statement to Resolve Belief-Inconsistency.

Relation Expressed By the Generality	Assumption Characteristics	Experiment 1 (Abstract)	Experiment 2 (Concrete)
Class	Combining	85.4 (5.2)	90.6 (7.8)
	Rending	57.5 (8.7) [a]	100.0 (1.1)
Property	Combining	72.3 (5.9)	80.8 (8.7)
	Rending	57.1 (11.9) [a]	78.6 (11.9)

[a] not reliably different from chance

Discussion

In prior studies of belief-conflict, when students were asked to reconcile inconsistencies among a set of *artificial relations*, their preference for retaining generalities on Combining problems ranged from 42% (Dieussaert, Schaeken, De Neys, & d'Ydewalle, 2000) to 70% (Revlis & Hayes, 1972). On Rending problems, the preference for retaining generalities ranged from as low as 25% (Dieussaert et al., 2000) to 47% (Revlin, Cate, & Rouss, 2001). In contrast, in the present study, where the artificial statements are thematically integrated, reasoners preferences are similar to those found for rich semantic categories (e.g., *All whales are mammals*) 91% and 61%, for

Combining and Rending, respectively (Revlin et al., 2001). This tells us that reasoners are capable of treating narrative-based class and property statements in the same way that they do beliefs about true semantic relations. We believe the important factor is that the arbitrary relations in the present study were integrated within the narrative.

The statements in each problem either described class-inclusion or property assignment. In the past, students have tended to treat generalities that expressed property assignment (*ducks have webbed feet*) as if they were less believable or law-like than generalities that expressed class-inclusion relations (*ducks are birds*) (e.g., Revlis & Hayes, 1972). However, participants in the present study, when reconciling inconsistencies among arbitrary relations tended to treat class and property statements equivalently. This may be a result of the underlying semantic relations used here: we have employed a context with binary distinctions in classes and properties (e.g., two kingdoms, two weapons, etc). Of course, the world hardly comes in such neat packages[1]. Future research will necessarily explore these findings in richer domains.

Experiment 2

How much power do physical situations exercise over our beliefs? Is seeing believing? An approach to this question adopted here is to use concrete materials about which students hold a belief and then notice how they revise their beliefs when confronted by a belief-contravening assumption.

Method

Participants. The participants were 55 student volunteers from the same class as participated in Experiment 1.

Materials & Procedures. The procedures employed in the present experiment are similar to that of Experiment 1 except that after reading each narrative, the participants were confronted with a diorama representing either of two locales (Fantasia and Gurge; or an American western town vs. a mountain pass). In addition, a Lego figure was used to represent a character from the narrative (a knight that either worked for King William or King Surly, etc.). The experimenter first read a set of statements from the problem (identical to a problem set from Experiment 1) and then pointed to the diorama and asked if the configuration of pieces were consistent with the story. Then the experimenter asked the participant to pretend that an assumption statement was true. This statement was equivalent to the appropriate belief-con-

---

[1] We thank Paolo Legrenzi for bringing this to our attention.

travening assumption from Experiment 1. The experimenter then asked "Do you want to leave things as they are or change things?"

The situation confronting the reasoner is illustrated with problem (1) above. The knight works for king William (1d) and has a battle-ax (1b). If the reasoner said to "leave things as they are", it would be equivalent to endorsing the particular statement; that is, the state of affairs that is presented as the physical condition, and reject the general statement that All of King Williams knights had swords (1a). If the reasoner selects to change things, that would be consistent with the general statement (knights have swords) and require the knight to have a sword instead of a battle-ax. This latter option is tantamount to retaining an abstract generalization and rejecting the situation "before the eyes" of the reasoner. (note: in an independent study, blank trials were included to estimate how often students want to change the Lego figure without there being any implications. Students spontaneously changed the figure less than 5% of the time). In front of the display, on a 7.5 cm. diameter circle were alternate horses, shields, and weapons. In nearly all cases where a change was made, the reasoner also changed all of the available properties of the figure (e.g., the color of the knight's weapon, its horse, etc.).

Results

If the reasoner said "change things", it was coded as supporting the generality. If the reasoner responded, "leaving things as they are", it was coded as rejecting the generality (and supporting the particular statement). Table 1 presents the percentage of trials on which reasoners retained the general statement. It reveals that the overall preference for the generality is 88%, ($p<.01$), which is reliably greater than would be expected by chance alone. Overall, this preference for generalities is equivalent for Combining and Rending problems and equivalent for affirmative and negative generalities. However, the preference to retain the generality is greater when they express class-inclusion than property-assignment relations [$F(1,41) = 4.4$, $p<.05$]. The effect of Relation does not interact with other variables.

Discussion

When students are given a choice between revising beliefs in story-based relations (generalities) or revising the physical characteristics of objects in front of them, that violate the story (particulars), they tend to reject the physical reality and make it conform to the abstract story relations.

Comparing across experiments (the students were all sampled from the same course) shows that overall, the preference for retaining generalities (and not revising them) is greater for concrete problems than for abstract

ones [$F(1,121)=9.0$, $p<.01$]. This is because on situations with a rending assumption, the reasoners tend to retain the generalities more often with concrete materials than with abstract ones [$F(1,20) = 9.9$, $p<.01$]. This in turn results in an interaction between the Type of Assumption (combining vs. rending) and the Concreteness (concrete vs. abstract) of the reasoning problem [$F(1,121)=3.8$, $p=.05$]. In sum, all abstract generalities are deemed more believable when the participants reason about physical objects.

General Discussion

We would like to highlight three aspects of this study for their contribution to our understanding of the belief revision process. First, this study shows that natural beliefs can be simulated by means of artificial relations that are woven together in a coherent narrative. This is not only a useful methodology for further research on belief revision, but offers us a first approximation to understanding the integrative nature of beliefs.

Second, once again we find that using the paper-and-pencil paradigm, when a belief-contravening assumption combines categories, reasoners tend to revise their beliefs so that they preserve the most general statements. In contrast, when a belief-contravening assumption divides-up, or rends, categories, reasoners show no particular preference to retain one type of statement over another.

Finally, the use of concrete materials in the present study opens up a new line of inquiry. Students reasoned about concrete materials as if they were fallible reflections of the category relations expressed in the story. Inconsistencies introduced by the belief-contravening assumption could have been resolved either by retaining the physical conditions that the reasoner perceived or by rejecting the facts-on-the-ground and thereby preserving the abstract knowledge structure presented in the story. It appears that reasoners chose the latter course and disregarded the facts before their eyes. This is reflected in the overwhelming preference for retaining the general statement as well as the absence of any difference in the preferences commonly shown for combining and rending assumptions. It has be proposed elsewhere (Revlin et al. 2001) that reasoners retain generalities less often on rending problems than on combining ones because the assumption challenges the immutability of the generality in the former situation. This was not found when physical objects were involved - it is the objects themselves that are considered mutable in the face of a belief-contravening assumption. Perhaps this is because one major use of counterfactuals in everyday life is to consider the consequences of objects and events being different from what they are: *what would happen if the dikes were not there?* As such, our participants may be well practiced in disregarding what they see.

The fact that reasoners treat the concrete situations as fallible is reminiscent of Plato's Allegory of the Cave in his *Republic*. In it, he proposes that our perceptions of the concrete world are based on distorted reflections of the ideal forms. The present data suggest that when people are forced to entertain belief-contravening assumptions about objects in the world, they readily relinquish the truth of what we see in favor of an organized abstraction. Perhaps in our everyday lives we tacitly apply Plato's allegory. As we pursue this line of research with concrete materials, we will be mindful of Vittorio Girotto's question (personal communication, July 22, 2004), *What has Plato wrought?*

References

Byrne, R. M. J. & Walsh, C. R. (2002). Contradictions and counterfactuals: Generating belief revisions in conditional inference. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Conference of the Cognitive Science Society*, 160-165.

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

Dieussaert, K., De Neys, W., & Schaeken, W. (2005). Suppression and belief revision: two sides of the same coin? *Psychologica Belgica, 45* (1), 29-46.

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

Ford, M. (2005). On using human nonmonotonic reasoning to inform artificial systems. *Psychologica Belgica, 45* (1), 57-70.

Legrenzi, P. & Johnson-Laird, P. N. (2005). The evaluation of diagnostic explanations for inconsistencies. *Psychologica Belgica, 45* (1), 19-28.

Nguyen, D. B. & Revlin, R. (1993). Transitive inferences from narrative relations. *Journal of Experimental Psychology: Learning, Memory & Cognition, 19*, 1197-1210.

Rescher, N. (1964). *Hypothetical reasoning*. Amsterdam: North-Holland Publishing.

Revlin, R., Cate, C. L., & Rouss, T. S. (2001). Reasoning counterfactually: Combining and rending. *Memory & Cognition, 29*, 1196-1208.

Revlis, R. & Hayes, J. R. M. (1972). The primacy of generalities in hypothetical reasoning. *Cognitive Psychology, 3*, 268-290.

Smith, E. E., Haviland, S. E., Buckley, P. B., & Sack, M. (1972). Retrieval of artificial facts from long-term memory. *Journal of Verbal Learning & Verbal Behavior, 11,* 583-593.

Wilkins, M. C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, *102*, 83.

Appendix

Fantasia Story

Many centuries ago there was a beautiful kingdom called Fantasia. The land in this kingdom was fertile and was filled with many crops that brought the kingdom great riches. Fantasia was ruled by a wise man, King William. Because King William was just and fair, his countrymen honored him by working hard to make the country flourish.

But the neighboring country of Gurge was not as well off as Fantasia. Gurge was ruled by a wicked and cruel man, King Surly, who robbed his countrymen of all their wealth. King Surly was so jealous of the good fortune of King William that he plotted to ruin Fantasia. King Surly decided to send a band of thieves into Fantasia to capture the king and burn the land. He went to his prison and selected the most evil prisoners to be in his band of thieves. He gave them each a battle-ax and a strong, black horse, and sent them on their way.

The band of thieves was very treacherous and began ruining many of the crops of Fantasia. King William was confused and upset until one of his spies told him of evil King Surly's plan. But King William knew how to defend his country. He called upon the bravest countrymen to be in his army. King William gave all of his soldiers powerful white horses and strong swords to protect themselves. The army was proud to defend their king. They raced through the country until they caught up with the evil band of thieves. The thieves were prepared to attack. But, the good king's army was relentless. The thieves began to grow weary. Many of them became frightened and fled. Soon the army had driven all the thieves away. After the battle, the army rode back to the castle where the king rewarded all the soldiers handsomely. The entire country then celebrated their victory.