Acta Psychologica 139 (2012) 486–491

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Hidden conflicts: Explanations make inconsistencies harder to detect☆

Sangeet S. Khemlani [a,*], P.N. Johnson-Laird [b]

[a] Naval Research Laboratory, United States
[b] Princeton University, United States

ARTICLE INFO

Article history:
Received 10 June 2011
Received in revised form 25 January 2012
Accepted 27 January 2012
Available online 22 February 2012

PsyINFO classification:
2340 Cognitive processes

Keywords:
Inconsistency
Explanations
Belief revision
Reasoning
Principle of resolution

ABSTRACT

A rational response to an inconsistent set of propositions is to revise it in a minimal way to restore consistency. A more important psychological goal is usually to create an explanation that resolves the inconsistency. We report five studies showing that once individuals have done so, they find inconsistencies harder to detect. Experiment 1 established the effect when participants explained inconsistencies, and Experiment 2 eliminated the possibility that the effect was a result of demand characteristics. Experiments 3a and 3b replicated the result, and showed that it did not occur in control groups that evaluated (or justified) which events in the pairs of assertions were more surprising. Experiment 4 replicated the previous findings, but the participants carried out all the conditions acting as their own controls. In all five studies, control conditions established that participants were able to detect comparable inconsistencies. Their explanations led them to re-interpret the generalizations as holding by default, and so they were less likely to treat the pairs of assertions as inconsistent. Explanations can accordingly undo the devastating consequences of logical inconsistencies, but at the cost of a subsequent failure to detect them.

1. Introduction

The word 'why' is used to elicit explanations for the mysteries of daily life. Why is my car making that noise? Why didn't the Redskins win last Sunday? Why isn't my experiment working? Indeed, a central feature of human rationality is the ability to construct explanations of observed phenomena (Harman, 1965). Recent research has explored the function and developmental trajectory of explanatory reasoning (Keil, 2006; Wellman, Hickling, & Schult, 1997). And there is consensus among researchers that explanations are related to causal inference (Johnson-Laird, Girotto, & Legrenzi, 2004; Sloman, 2005; Byrne, 2005; Walsh & Johnson-Laird, 2009), and that explanations affect reasoning, categorization, and learning (Lombrozo, 2006). Less is known about the contexts in which individuals create explanations, i.e., when and how they decide to produce explanations. One obvious context is when they are asked for an explanation. But, people also produce explanations when they are learning new information (Amsterlaw & Wellman, 2006; Chi, De Leeuw, Chiu, & Lavancher, 1994; Crowley & Siegler, 1999; Rittle-Johnson, 2006), trying to form categories (Shafto & Coley, 2003), and judging how well concepts cohere with one another (Murphy & Medin, 1985; Patalano, Chin-Parker, & Ross, 2006). Explanations also help individuals to predict future behaviors (Anderson & Ross, 1980; Einhorn & Hogarth, 1986; Lombrozo & Carey, 2006; Ross, Lepper, Strack, & Steinmetz, 1977).

Another context in which individuals spontaneously create explanations is when they detect inconsistencies (Johnson-Laird, 2006; Johnson-Laird et al., 2004; Khemlani & Johnson-Laird, 2011). Even children are likely to generate causal explanations if they observe an inconsistency with their previous experience in an experiment (Legare, 2012; Legare, Gelman, & Wellman, 2010). The relation between inconsistency and explanation is the topic of the present paper, and, in particular, how explanations can in turn make inconsistencies harder to detect.

2. Explanations resolve inconsistencies

When individuals detect an inconsistency among a set of assertions, they try to explain the origins of the inconsistency. If they know what created the inconsistency then they can make a better decision about an appropriate course of action. The explanation, of course, has a side effect of restoring consistency to the set of propositions (Johnson-Laird et al., 2004; Khemlani & Johnson-Laird, 2011). According to this *principle of resolution*, they then re-interpret the inconsistent assertions based on the consequences of their explanation.

☆ This research was supported by a National Science Foundation Graduate Research Fellowship to the first author, and by National Science Foundation Grant no. SES 0844851 to the second author to study deductive and probabilistic reasoning. For their helpful comments and criticisms, we thank Jeremy Boyd, John Darley, Sam Glucksberg, Adele Goldberg, Geoffrey Goodwin, Matt Johnson, Olivia Kang, Niklas Kunze, Max Lotstein, Joe Magliano, Gorka Navarrete, Ray Nickerson, Danny Oppenheimer, Laura Suttle, Rolf Zwaan, and the audience at the 32nd Annual Conference of the Cognitive Science Society, where an earlier version of this paper was presented.

* Corresponding author at: Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, Washington, DC 20375, United States. Tel.: +1 5713399312.

E-mail address: skhemlani@gmail.com (S.S. Khemlani).

0001-6918/$ – see front matter
doi:10.1016/j.actpsy.2012.01.010

Consider, for instance, the following scenario:

> If a person does regular aerobic exercises then the person strengthens her heart.
>
> Someone did regular aerobic exercises, but she did not strengthen her heart.

The two assertions are inconsistent, i.e., they cannot both be true. Given such an inconsistency, it is felicitous to ask: “why not?” One explanation for the person failing to strengthen her heart is that she had a heart defect. This explanation provides an exception to the generalization about regular aerobic exercise. It also suggests that an appropriate course of action is to seek a cure for the heart defect. Individuals could abandon the generalization as false, or, more likely, construe it as an idealization that holds by default: it is true in typical cases, but it tolerates exceptions. The assertion is accordingly interpreted as akin to the *generic* assertion, i.e., aerobic exercises strengthen the heart, which also tolerates exceptions (Leslie, 2008; Leslie, Khemlani, & Glucksberg, 2011). The principle of resolution therefore predicts that individuals create explanations to resolve inconsistencies, and that the explanations can lead to the tacit re-interpretation of general assertions as generics that hold by default. This re-interpretation yields a prediction: once individuals have formulated an explanation that resolves an inconsistency, they should be less likely to detect the inconsistency. An alternative hypothesis is that explanations have no effect on the status of inconsistencies, and so the ability to detect them is not subsequently impaired. To test these contrasting predictions, we carried out five experiments in which participants had to detect inconsistencies before or after they constructed explanations of them.

3. Experiment 1

According to the principle of resolution, explanations lead to re-interpretations of inconsistent assertions, and as a result an interaction should occur: inconsistency should be harder to evaluate after individuals have explained what's going on than beforehand, and this effect should be greater than the effect of explanations on judgments of consistency. Experiment 1 tested this prediction. The participants were presented with pairs of assertions, such as:

> If a person is bitten by a viper then the person dies.
>
> Someone was bitten by a viper, but did not die.

They also answered the question, “why not?” either before or after they evaluated the consistency of the assertions. The question, of course, called for them to explain what is going on.

3.1. Method

3.1.1. Participants

40 participants volunteered through the Mechanical Turk online platform hosted through Amazon.com. Mechanical Turk is a system that distributes tasks, surveys, and experiments to thousands of people for completion. The platform is a viable alternative to laboratory experimentation (see Paolacci, Chandler, & Ipeirotis, 2010, for a comparison of different recruitment methods). Participants volunteered for the study through a listing of studies, and they completed it for monetary compensation (in the form of credit towards their accounts). They could complete the study only once, and the pool of participants was constrained to meet several demographic specifications. The experiment was made available to a) only North American Amazon.com subscribers, b) only those participants who self-reported that they were native English speakers, and c) only those participants who reported that they had no prior expertise in logic.

3.1.2. Design and procedure

On each trial, participants were given a pair of consistent or inconsistent assertions. There were two different groups: 20 participants performed an explanation task before they evaluated the consistency of pairs of assertions, and 20 performed the two tasks in the opposite order. For both groups, half of the problems contained a generalization (1) that was inconsistent with a categorical assertion (2), e.g.,

1. If someone is very kind then he or she is liked by others.
2. Someone was very kind but was not liked by others.

For the other half of the problems, the inconsistency was eliminated by omitting the first clause in the categorical assertion, e.g.,

3. If someone is very kind then he or she is liked by others.
4. Someone was not liked by others.

Participants received equal numbers of consistent and inconsistent problems, and carried out the two tasks in succession for each problem. For the consistency task, they had to answer the question, “Can both of these statements be true at the same time?” (We used this question because participants are often uncertain about the meaning of “consistent” whereas the question is unambiguous.) They responded by pressing one of two buttons marked “Yes” or “No”. For the explanatory task, they answered the question, “Why not?” The question made sense for both the consistent and inconsistent pairs, because the final clause in both sorts of problem was a negative assertion. The participants typed their answers into a text box provided on the screen. They were unable to see their answer to the first task when they carried out the second task. All of the problems were similar to the two examples above (see the Appendix A for all the materials in the experiments). Each participant encountered a given pair of assertions only once, and received the pairs in a different random order. The participants were given no clues that the generalizations in the study had exceptions. Instead, they were told that the experiment was about conflicts in information, and that they would have to carry out both an evaluation task and an explanatory task. The two tasks were illustrated for them as follows:

Suppose you are told the following:

1. *If a food item is not preserved, then it rots.*
2. *This food item was not preserved, but it did not rot.*

Based on this information, you will be asked to explain what is going on. For every trial, you will also be asked if both sentences can be true at the same time. Please respond based on what you think the appropriate answer is.

3.2. Results and discussion

Table 1 presents the percentages of correct evaluations in Experiment 1. Participants were far more accurate at detecting consistencies than inconsistencies (89% vs. 45%, Wilcoxon test, $z = 4.00$, $p < .0001$, Cliff's $\delta = .69$). The group that evaluated the consistency of the assertions first was more accurate than the group that provided an explanation first (79% vs. 56%, Mann–Whitney test, $z = 3.07$, $p < .005$, Cliff's $\delta = .66$). Likewise, the predicted interaction was significant: the difference in accuracy between consistent and inconsistent problems

Table 1
The percentages of correct evaluations of consistency and inconsistency in Experiment 1 depending on whether this task occurred before or after the explanatory task.

	Inconsistent problems	Consistent problems
Group that carried out the consistency task first	64	93
Group that carried out the explanatory task first	27	86