Negativity bias in defeasible reasoning

Lupita Estefania Gazzo Castañeda, Bruno Richter & Markus Knauff

To cite this article: Lupita Estefania Gazzo Castañeda, Bruno Richter & Markus Knauff (2016) Negativity bias in defeasible reasoning, Thinking & Reasoning, 22:2, 209-220, DOI: 10.1080/13546783.2015.1117988

To link to this article: http://dx.doi.org/10.1080/13546783.2015.1117988

Published online: 09 Dec 2015.

Article views: 222

View related articles

View Crossmark data
Negativity bias in defeasible reasoning

Lupita Estefania Gazzo Castañeda, Bruno Richter and Markus Knauff

Experimental Psychology and Cognitive Science, Department of Psychology, University of Giessen, Giessen, Germany

ABSTRACT

In defeasible reasoning, initially drawn conclusions can be withdrawn in light of new information. In this paper, we examine how the conclusions drawn from conditionals describing positive or negative situations can be defeated by subsequent negative or positive information, respectively. Participants were confronted with conditionals of the form “If [situation], then I am happy/sad” which were either followed by no additional information or by additional information describing situations of the same or the opposite valence. The participant’s task was to decide on a question asking for a possible conclusion (“Am I happy?” vs. “Am I sad?”). We found a negativity bias in defeasible reasoning: negative information defeated positively charged conclusions more strongly than positive information defeated negatively charged conclusions. We discuss our results in relation to the new psychology of reasoning.

ARTICLE HISTORY

Received 25 February 2015; Accepted 2 November 2015

KEYWORDS

Conditionals; defeasible reasoning; reasoning; negativity bias

Introduction

Consider the following problem:

If my mother dies, then I am sad.
My mother dies.
I get my dream job.
Am I sad?

How would you answer? Would you conclude being sad even though you get your dream job? Now consider this information:

If I get my dream job, then I am happy.
I get my dream job.
My mother dies.
Am I happy?

Now how would you answer? Would you conclude being happy even though your mother died?

CONTACT

Lupita Estefania Gazzo Castañeda Estefania.Gazzo@psychol.uni-giessen.de

© 2015 Taylor & Francis
According to classical logic, the valid conclusion in both problems is “yes”. In both cases, the reasoner is presented with a conditional “if-then” rule together with the fact that the if-part (i.e., the antecedent) of the rule is given so that he or she can infer logically that the when-part (i.e., the consequent) of the rule follows. This rule is called the Modus Ponens inference. However, recent research in cognitive psychology has shown that people often defeat valid conclusions if they can think of — or are confronted with — information which is perceived to prevent the consequent of the conditional to occur although the antecedent is given (e.g., Byrne, 1989; Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys, Schaeeken, & d’Ydewalle 2003a, 2003b; Dieussaert, De Neys, & Schaeeken, 2005; Johnson-Laird & Byrne, 2002; Weidenfeld, Oberauer, & Hörning, 2005; see also Oaksford & Chater, 2001; 2003a; Oaksford, Chater, Larkin, 2000). In the first example, people might withdraw from concluding being sad when faced with the fact that something positive has happened, like getting one’s dream job. The second example is the same in the opposite direction: people might withdraw from concluding being happy when faced with the fact that something negative has happened, like one’s mother dying. This withdrawal of otherwise valid conclusions in light of additional information is called defeasible reasoning (e.g., Evans, 2002; Oaksford & Chater, 1995; Politzer, 2007). Literature on defeasible reasoning has shown that the more exceptions to a conclusion a person considers, the less he or she will follow the initial conditional rule and the more this person will withdraw the otherwise valid conclusions (De Neys et al., 2003a, 2003b). But is every potential exception also equally strong in defeating a conclusion? It has been argued that exceptions differ in their associative strength or relative salience (Chan & Chua, 1994; De Neys, Schaeeken, & d’Ydewalle 2002; De Neys et al., 2003b; see also Markovits & Potvin, 2001). The probability of accepting an exception is thought to depend on how strongly it is associated to the consequent of the conditional in semantic memory (De Neys et al., 2003b; see also Markovits, Fleury, Quinn, & Venet, 1998; Quinn & Markovits, 1998). In this way, for a conditional such as “If the apples are ripe, then they fall from the tree”, the information of the apples being picked has more defeasible power than the information of the apples being caught in the branches (see De Neys et al., 2003b). However, although the semantic memory approach works for many conditionals, we think that such an explanation is not sufficient for explaining the defeasible power of exceptions when reasoning with emotionally charged conditionals like the ones in the beginning of this paper. Can positively charged information (in the following: positive information) defeat initial negatively charged conclusions as strongly as negatively charged information (in the following: negative information) can defeat initial positively charged conclusions?

Psychological research on attitudes and emotions has shown the existence of a so-called negativity bias (see Cacioppo, Gardner, & Berntson, 1997; Peeters & Czapisnki, 1990). People “respond more strongly to very negative
stimuli than to matched positive stimuli” (Norris, Larsen, Crawford, & Cacioppo, 2011, p. 100; see also Ito, Cacioppo, & Lang, 1998). As Vaish, Grossman and Woodward (2008) review, in comparison to positive information, negative information is observed longer (e.g., Fiske, 1980), weight more heavily when making decisions and evaluations (e.g., Atthowe 1960; Kahneman & Tversky, 1983; Kanouse & Hanson, 1972), and considered more when making inferences about traits of other people (e.g., Aloise, 1993; Skowronski & Carlston, 1989; Wyer & Hinkle, 1976; for reviews on the negativity bias, see Peeters & Czapinski, 1990; Vaish et al., 2008). The main explanation for the negativity bias is evolutionary. It is argued that the negativity bias helps organisms to recognize danger and avoid hostile environments (Cacioppo et al., 1997; Cacioppo, Gardner, & Berntson, 1999; Peeters & Czapinski, 1990; Vaish et al., 2008). Negative information, therefore, is more informative than positive information (Peeters & Czapinski, 1990; Vaish et al., 2008).

The existence of the negativity bias suggests that negative information has more defeasible power than positive information. It should thus be easier to defeat a positively charged conclusion by additional negative information than defeating an initially negatively charged conclusion by positive information. Applying this to the initial examples, participants should withdraw the second example’s conclusion more readily than the conclusion in the first example. We tested this hypothesis by presenting participants conditional inferences which either had no additional information or additional information evoking the same or the opposite emotion described in the conditional. The amount of defeated conclusions is used as dependent measure for a negativity bias in defeasible conditional reasoning.

**Methods**

**Participants**

Thirty participants (15 female) took part in the experiment. All but one participant were students. The mean age was $M = 24.83$ years ($SD = 3.82$).

**Material and design**

We constructed 36 conditional inference problems, 24 of them with additional information like the examples in the beginning of this paper (i.e., experimental items), and the remaining 12 problems without any additional information (i.e., control items; for a similar design, see Hilton, Jaspars, & Clarke, 1990). The problems therefore consisted of (1) an initial conditional rule, (2) the fact that the antecedent of this rule is given, (3) an additional information (omitted for control items), and (4) the question about the conclusion. We experimentally varied the valence of the initial conditional rule and the valence of the additional information by using situations rated as
positive or negative by $N = 255$ participants in an online preliminary study via SoSci Survey (Leiner, 2014). In this preliminary study, participants had to rate how they feel about $n = 31$ positive and $n = 31$ negative situations on a 7-point Likert scale ($1 = \text{very sad}$, $7 = \text{very happy}$). For the experiment, we selected the 6 most positive ($M = 6.40; SD = 0.16$) and the 6 most negative ($M = 1.50; SD = 0.26$) situations and brought them into a conditional form (i.e., “If [the situation], then I am sad/happy.”). Each conditional rule was presented three times, once with additional information of the same valence, once with additional information of the opposite valance, and once without any additional information. As additional information, we used the same six positive and six negative situations we used to create the conditionals but without framing them into a conditional form. Hereby, we alternated which situation was presented as the conditional rule and which information was presented as the additional information. As illustrated in the two examples in the beginning of this paper, in half of the cases, a specific situation $A$ was phrased as the conditional rule, together with situation $B$ as the additional information. In the other half of the cases, however, it was the other way around: the situation $B$ was now presented as the conditional rule, and situation $A$ as the additional information. We made sure that the pairs of positive and negative situations used in one inference problem were matched in strength, i.e., ratings for positive and negative situations were equally distant from the neutral scale midpoint. In a second step, in order to guarantee that the pairs of situations in each inference problem were not only matched in the strength of their emotions, but also in their associative strength, we conducted a second validation study. Via SoSci Survey (Leiner, 2014) we presented each of the 12 situations (and 12 filler situations) in a conditional form paired once with the positive and once with the negative consequent (e.g., “If your mother dies, are you then sad?”). Participants ($N = 67$) had to answer on a 7-point Likert scale ranging from $1 = \text{definitely not}$ to $7 = \text{definitely yes}$. Results corroborated that positive situations were as strongly associated to the “I am happy” consequents ($M = 6.67; SD = 0.40$) as were the negative situations to the “I am sad” consequents ($M = 6.56; SD = 0.47$), $t(66) = 1.63$, $p = .109, d = 0.23^{1}$.

A complete list of the items used can be found in the table of the Appendix. The experiment followed a 2 (valence of the conditional: positive vs. negative) $\times$ 3 (valence of the additional information: positive vs. negative vs. none) within subjects design.

**Procedure**

The experiment was conducted with Cedrus SuperLab © 4.5 on a desktop computer. Participants were tested individually and instructed to answer spontaneously to the question about the conclusion in the end of each...
problem. They were told that there exists no right or wrong answer. We did not ask participants to assume the premises as true and to answer according to logical necessity, because we were interested in the natural understanding of conditionals (cf Cummins, 1995; De Neys et al., 2003a, 2003b). Participants gave their answers by either pressing a “Y” (yes) or a “N” (no) key on the keyboard. The order of these keys was counterbalanced. The spacebar was used to switch from one premise to the next one. The question at the conclusion was written in red, the premises in black. After each problem, participants had the possibility to take a break. Before starting the experiment, participants completed two practice problems.

**Results**

We computed the amount of defeated conclusions in per cent for each category of problems separately. One participant had to be excluded from our computations because he or she deviated over 4 SD from the mean in two of the four experimental conditions with additional information. Descriptive results can be found in Figure 1.

**Control problems with no additional information**

An important prerequisite for our experiment was that when no additional information is presented, participants make the logically valid conclusion.
Therefore, we analysed how often participants defeated the logically valid conclusion of control problems. Without any additional information, participants almost never defeated the conclusion suggested by the conditional rule, neither for positive ($M = 0.0\%$; $SD = 0.0$) nor for negative ($M = 3.45\%$, $SD = 11.25$) conditionals. In fact, a $t$-test for repeated measures shows that the participants’ acceptance of positively charged and negatively charged conclusions did not differ from each other, $t(28) = 1.65$, $p = .110$, $d = 0.43$.

### Problems with additional positive or negative information

To test the impact of additional information, we conducted a $2$ (valence of the conditional: positive vs. negative) $\times 2$ (valence of the additional information: positive vs. negative) analysis of variance (ANOVA) on the percentages of defeated conclusions of the experimental problems. As expected, the ANOVA showed a main effect for the valence of the conditional, $F(1, 28) = 20.64$, $p < .001$, $\eta_p^2 = .424$, a main effect for the valence of the additional information, $F(1, 28) = 41.90$, $p < .001$, $\eta_p^2 = .599$, and an interaction between both factors, $F(1, 28) = 204.79$, $p < .001$, $\eta_p^2 = .880$. We analysed this interaction with post hoc $t$-tests and a Bonferroni adjusted alpha level of 0.025. As can be seen in Figure 1, positively charged conclusions were defeated by negative additional information ($M = 74.71\%$; $SD = 28.04$) more often than were negatively charged conclusions by positive additional information ($M = 33.33\%$; $SD = 27.46$), $t(28) = 5.64$, $p < .001$, $d = 1.49$. When the conditional rule was paired with additional information of the same valence, participants did not differ in the percentage of defeated conclusions ($M_{\text{neg-neg}} = 4.02\%$; $SD = 11.49$; $M_{\text{pos-pos}} = 0.57\%$; $SD = 3.09$), $t(28) = 1.54$, $p = .136$, $d = 0.42$.

### Discussion

In this study, we were interested in how positively charged and negatively charged additional information can defeat negative and positive conclusions, respectively. Our results show that negative additional information has more power in defeating positively charged conclusions than positive information has in defeating negatively charged conclusions. Since in this experiment positive and negative information were of the same strength, the mere associative strength between additional information and their respective conclusions of being sad or being happy cannot explain these results. Instead, the higher defeasible power of negative information can be seen as a negativity bias in defeasible reasoning. As already shown for decision-making (e.g., Atthowe 1960; Kahneman & Tversky, 1983) and impression formation (e.g., Aloise, 1993; Skowronski & Carlston, 1989; Wyer & Hinkle, 1976), people also seem to respond stronger to negative information compared to positive
information in defeasible reasoning. This negativity bias makes people to be more influenced by negative than by positive information when trying to defeat a conclusion of the opposed valence.

One possible reason for the superiority of negative compared to positive information is that negative information can be more diagnostic (e.g., Skowronski and Carlston, 1987, 1989; for an overview, see Lewicka, Czapinski, & Peeters, 1992), or less expected than positive information (Ikegami, 1993). Another explanation provides the evaluative space model, which explains the negativity bias by assuming two separable underlying motivational systems functioning at different levels of evaluative input (Cacioppo & Berntson, 1999; Cacioppo et al., 1997, 1999; Ito & Cacioppo, 2005). Further studies can help in understanding which mechanisms explain the negativity bias we found in our defeasible reasoning paradigm.

Besides supporting previous findings on the negativity bias, our findings also have implications for cognitive psychology. In particular, our study shows that human reasoning is defeasible and influenced by everyday experiences. This agrees with the main assumptions of the so-called “new psychology of reasoning” (Evans, 2012). Several studies have shown that background knowledge (e.g., Chan & Chua, 1994; Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003a, 2003b; Johnson-Laird & Byrne, 2002), expected utilities (e.g., Bonnefon, 2009; Bonnefon, Haigh, & Steward, 2013; Manktelow & Over, 1991), probabilities (e.g., Evans, Handley, & Over, 2003; Oaksford & Chater, 2003b, 2007; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007), and more indirect factors like the reasoner’s personality (Bonnefon, 2010), and societal rules (Demeure, Bonnefon, & Raufaste, 2009) affect the acceptability of conclusions. In our study, we extend these findings by showing that also the emotional valence of conditionals and additional information affects defeasible reasoning. Given that we kept associative strength constant, the different weightings people give to positive and negative information cannot be easily explained by current theories of reasoning, such as mental model theory (Johnson-Laird & Byrne, 1991, 2002) or probability-based approaches (e.g., Evans & Over, 2004; Oaksford & Chater, 2007). To explain our results, it is necessary to assume either something like comparative mental models which weight some premises higher than others depending on which other premises they are paired with, or some kind of probability computation which prescribes how to weight conflicting premises with equal conditional probabilities.

We are aware that the role of emotions on reasoning has already received attention in the reasoning literature (e.g., Blanchette, 2006; Blanchette & Leese, 2010; Jung, Wranke, Hamburger, & Knauff, 2014; for a review, see Blanchette, 2014). However, those studies were mainly interested in how emotions influence classical, deductive reasoning, showing, e.g., that strong
emotions lower reasoning performance. Our study is different because we were concerned with how emotions influence the weighting of additional information. In fact, we know of only two studies that investigated how the emotions evoked by additional information influence conclusion endorsement. For instance, Bonnefon and Hilton (2004) showed that when a conditional, such as “If Marie’s TV is broken, she will have it fixed” is followed by “If Marie has her TV fixed, she will not be able to pay the electricity bill”, people refuse to endorse the MP inference that Marie will have the TV fixed. Yet, Bonnefon and Hilton (2004) worked with consequential conditionals where people have to decide whether an action described in the conditional will be taken or not. Our conditionals instead describe events beyond the agent’s control and reasoners have to decide which emotion prevails. Also Verschueren, Peeters, and Schaeken (2006) showed that people make more MP conclusions and think less of exceptions when a conditional is of positive than of negative valence. In this way, Verschueren et al. (2006) already pointed to some kind of positive—negative asymmetry in conditional reasoning. However, our study expands these findings because we show how conflicting emotions are weighted.

For further studies, it would be interesting to test for individual differences. Individual differences in the negativity bias are known (Ashare, Norris, Wileyto, Cacioppo, & Strasser, 2013; Ito & Cacioppo, 2005; Norris et al., 2011) and we may expect that these individual differences also influence the weight people give to positive and negative additional information during reasoning. For instance, optimists might give more weight to positive information than pessimists, and pessimists might give more weight to negative information than optimist (cf. Norris et al., 2011). We are currently running experiments on this and hope to improve the understanding of how human beings reason with conditionals and which factors influence the conclusions they draw—or withdraw.

Note

1. Standardised mean differences (d) were computed as described by Borenstein (2009).

Acknowledgements

We thank Denis Hilton and two anonymous reviewers for their helpful comments and suggestions on earlier versions of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.
Funding

This research was supported by DFG [grant KN 465/10-1 and KN 465/10-2] within the Priority Program “New Frameworks of Rationality” (SPP 1516) to Markus Knauff.

References


## Appendix

Items used in the experiment.

<table>
<thead>
<tr>
<th>Negatively charged conclusion</th>
<th>Positively charged conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R</strong>: If my mother dies, then I am sad.</td>
<td><strong>R</strong>: If I get my dream job, then I am happy.</td>
</tr>
<tr>
<td><strong>F</strong>: My mother dies.</td>
<td><strong>F</strong>: I get my dream job.</td>
</tr>
<tr>
<td><strong>Al</strong>: My flat is set on fire. [OV]</td>
<td><strong>Al</strong>: My partner tells me that (s)he loves me. [SV] [none]</td>
</tr>
<tr>
<td>My mother dies. [SV] [none]</td>
<td><strong>C</strong>: Am I happy?</td>
</tr>
<tr>
<td><strong>C</strong>: Am I sad?</td>
<td></td>
</tr>
<tr>
<td><strong>R</strong>: If my flat is set on fire, then I am sad.</td>
<td><strong>R</strong>: If I discover that my partner is cheating on me, then I am sad.</td>
</tr>
<tr>
<td><strong>F</strong>: My flat is set on fire.</td>
<td><strong>F</strong>: I discover that my partner is cheating on me.</td>
</tr>
<tr>
<td><strong>Al</strong>: My partner tells me that (s)he loves me. [OV]</td>
<td><strong>Al</strong>: I passed my final examination successfully. [SV] [none]</td>
</tr>
<tr>
<td>My mother dies. [SV] [none]</td>
<td><strong>C</strong>: Am I happy?</td>
</tr>
<tr>
<td><strong>C</strong>: Am I sad?</td>
<td></td>
</tr>
<tr>
<td><strong>R</strong>: If I discover that my partner is cheating on me, then I am sad.</td>
<td><strong>R</strong>: If I reach a goal I pursued for a long time, then I am happy.</td>
</tr>
<tr>
<td><strong>F</strong>: I discover that my partner is cheating on me.</td>
<td><strong>F</strong>: I reach a goal I pursued for a long time.</td>
</tr>
<tr>
<td><strong>Al</strong>: I passed my final examination successfully. [OV]</td>
<td><strong>Al</strong>: I discovered that my partner is cheating on me. [OV]</td>
</tr>
<tr>
<td>My partner breaks up with me. [SV]</td>
<td><strong>C</strong>: Am I happy?</td>
</tr>
<tr>
<td>[none]</td>
<td></td>
</tr>
<tr>
<td><strong>C</strong>: Am I sad?</td>
<td></td>
</tr>
<tr>
<td><strong>R</strong>: If my partner breaks up with me, then I am sad.</td>
<td><strong>R</strong>: If I lose my job, then I am sad.</td>
</tr>
<tr>
<td><strong>F</strong>: My partner breaks up with me.</td>
<td><strong>F</strong>: I lose my job.</td>
</tr>
<tr>
<td><strong>Al</strong>: I reach a goal I pursued for a long time. [OV]</td>
<td><strong>Al</strong>: My pet dies. [OV]</td>
</tr>
<tr>
<td>I discover that my partner is cheating on me. [SV] [none]</td>
<td><strong>C</strong>: Am I happy?</td>
</tr>
<tr>
<td>[none]</td>
<td></td>
</tr>
<tr>
<td><strong>C</strong>: Am I sad?</td>
<td></td>
</tr>
<tr>
<td><strong>R</strong>: If I lose my job, then I am sad.</td>
<td><strong>R</strong>: If I laugh lustily with a good friend, then I am happy.</td>
</tr>
<tr>
<td><strong>F</strong>: I lose my job.</td>
<td><strong>F</strong>: I laugh lustily with a good friend.</td>
</tr>
<tr>
<td><strong>Al</strong>: I see my family after a long time. [OV]</td>
<td><strong>Al</strong>: I lose my job. [OV]</td>
</tr>
<tr>
<td>I lose my job. [SV] [none]</td>
<td><strong>C</strong>: Am I happy?</td>
</tr>
<tr>
<td>[none]</td>
<td></td>
</tr>
<tr>
<td><strong>C</strong>: Am I sad?</td>
<td></td>
</tr>
<tr>
<td><strong>R</strong>: If my pet dies, then I am sad.</td>
<td><strong>R</strong>: If I see my family after a long time, then I am happy.</td>
</tr>
<tr>
<td><strong>F</strong>: My pet dies.</td>
<td><strong>F</strong>: I see my family after a long time.</td>
</tr>
<tr>
<td><strong>Al</strong>: I lose my job.</td>
<td><strong>Al</strong>: My pet dies. [OV]</td>
</tr>
<tr>
<td>I see my family after a long time. [OV]</td>
<td><strong>C</strong>: Am I happy?</td>
</tr>
<tr>
<td>[none]</td>
<td></td>
</tr>
<tr>
<td><strong>C</strong>: Am I sad?</td>
<td></td>
</tr>
</tbody>
</table>

Note: Each rule was presented three times: once with additional information of the opposite valence (OV), once with additional information of the same valence (SV), and once without any additional information (none). **R** = rule; **F** = fact; **Al** = additional information; **C** = conclusion.