Thinking & Reasoning

ISSN: 1354-6783 (Print) 1464-0708 (Online) Journal homepage: https://www.tandfonline.com/loi/ptar20

The specificity of terms affects conditional reasoning

Lupita Estefania Gazzo Castañeda & Markus Knauff

THINKING & REASONING
2019, VOL. 25, NO. 1, 72-93
https://doi.org/10.1080/13546783.2018.1467343

Routledge
Taylor & Francis Group

Check for updates

The specificity of terms affects conditional reasoning

Lupita Estefania Gazzo Castañeda and Markus Knauff

Experimental Psychology and Cognitive Science, Justus Liebig University Giessen, Giessen, Germany

ABSTRACT
Conditional inferences can be phrased with unspecific terms ("If a person is on a diet, then the person loses weight. A person is on a diet. The person loses weight") or specific terms ("If Anna is on a diet, then Anna loses weight. Anna is on a diet. Anna loses weight"). We investigate whether the specificity of terms affects people's acceptance of inferences. In Experiment 1, inferences with specific terms received higher acceptance ratings than inferences with unspecific terms. In Experiments 2 and 3, we used the same problems as in Experiment 1 but also problems with unspecific terms in the conditional and specific terms in the categorical and vice versa. When the conditional and the categorical had the same specificity, results were as in Experiment 1. When the specificity of the conditional and the categorical mismatched, acceptance ratings were lower. Our results illustrate the importance of phrasing on reasoning.

ARTICLE HISTORY Received 10 April 2017; Accepted 16 April 2018

KEYWORDS Reasoning; conditionals; specificity; phrasing; pragmatics

Consider the following inferences:

If a person studies hard, then the person will do well on the test.
A person studies hard.

The person will do well on the test.
vs.
If Anna studies hard, then Anna will do well on the test.
Anna studies hard.

Anna will do well on the test.

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

Which inference would you accept more? Both inferences have the same structure and both are about the same topic. Therefore, one could expect that both are accepted into the same degree. However, in this paper, we argue that the second inference should be accepted more than the first one. We now explain why.

According to classical logic, both inferences are equally valid. We have a *conditional premise* "if *p*, then *q*", a *categorical premise* containing *p* and a *conclusion* stating *q*. According to classical logic, this conclusion is true because the logical connective "if" implies that if *p* is true (here: studying hard), then *q* is also true (here: doing well on the test) – a valid inference called *modus ponens* (MP). In a similar way, also the *modus tollens* (MT) inference is valid (if *p* then *q*; not-*q*; therefore not-*p*). However, contrary to classical logic, people sometimes refuse to draw such valid inferences. Further, they often even consider invalid inferences as valid, as for example the inferences *affirmation of the consequence* (AC: if *p*, then *q*; *q*; therefore *p*) and *denial of the antecedent* (DA: if *p*, then *q*; not-*p*; therefore not-*q*). Both are invalid, because in classical logic *p* is only sufficient, but not necessary for *q* (e.g., Hilton, Jaspars, & Clarke, 1990).

One important reason for this pattern of results is that people do not only reason according to the structure of a task. Instead, they also consider the content of a conditional (see e.g., De Neys, Schaeken, & d'Ydewalle, 2003a, 2003b; Evans & Over, 2004; Johnson-Laird & Byrne, 2002; Oaksford & Chater, 2007). For instance, in the initial example, the conditional contains the information that studying hard leads to a good result in the test. This might be generally true. However, we know that there are circumstances that prevent this: blackouts, little sleep or a too hard test can prevent one from doing well on tests. When people consider these circumstances during reasoning, their willingness to accept inferences decreases (e.g., Cummins, 1995; De Neys et al., 2003a). Here, it is important to distinguish between *disablers* and *alternatives*. Both affect people's acceptance of conclusions. Disablers are circumstances that prevent *q* from happening although *p* is the case, such as a too hard test, blackouts or little sleep in our initial example. The more disablers a person considers, the less MP and MT conclusions are drawn (Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys et al., 2003a). Alternatives are circumstances that bring about *q*, but which are not *p*. In our initial example, such alternatives could be a very easy test or using a cheat slip. For instance, when reasoners consider the possibility that the test might be very easy, then their willingness for concluding that *p* (here: studying hard) is necessary for *q* (here: doing well on the test) decreases. Therefore, the more alternatives a person considers, the less AC and DA inferences are drawn (Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003a). In other words, disablers and alternatives affect the sufficiency and necessity relation between *p* and *q* (e.g., Thompson, 1994, 1995). As a result, many disablers also lead to a low likelihood of *q* given *p*, and many alternatives to a low likelihood of *p* given *q* (Verschueren, Schaeken, & d'Ydewalle, 2005; see also Evans & Over, 2004; Thompson, 1995).

However, to be able to consider disablers and alternatives, it is necessary to have knowledge about the content of the conditional. Cummins (1995) showed that participants accept more valid and invalid inferences when conditionals have fictitious content (e.g., "If it thardrons, then the streets will be sticky") instead of everyday content (e.g., "If it rains, then the streets will be wet"). Markovits (1986) showed that the acceptance of DA and AC inferences was higher for unfamiliar content (e.g., "If an augmented chord is played, a three-tone harmonic interval is produced") than for familiar content (e.g., "If one eats poisoned mushrooms, one becomes ill"). Chan and Chua (1994) showed that people can fail to weigh disablers correctly when confronted with domain-specific conditionals they do not know much about. Gazzo Castañeda and Knauff (2016a) showed that the consideration of disablers in legal reasoning is lower for laypeople than for lawyers. All these findings show that lack of knowledge leads to higher acceptance of inferences: when people have no knowledge about the content of a conditional, then they also have no knowledge about possible disablers or alternatives that can affect inferences (Cummins, 1995).

Another factor that affects people's consideration of background knowledge is the phrasing of conditionals. In Gazzo Castañeda and Knauff (2017), we re-phrased MP and MT inferences as quantified statements and found that people's consideration of disablers depended on the quantifier. When a conditional was re-phrased as a universal statement ("All persons that study hard will do well in tests"), then participants considered less their knowledge about disablers than when the conditional was re-phrased as an existential statement ("Some persons that study hard will do well in tests"). We explained this effect of quantifier by arguing that different phrasings have different pragmatics: while universal quantifiers negate the existence of disablers, existential quantifiers suggest that disablers do exist (Gazzo Castañeda & Knauff, 2017). In a similar way, in Gazzo Castañeda and Knauff (2016b), we found that the phrasing of modal auxiliary also affects inferences. When the conditional contained the modal auxiliary *should* ("If a person downloads child pornography, then the person should be punished for possession of child pornography"), then participants made more MP inferences concluding that the offender should be punished – ignoring possible disablers. But when the conditional contained the modal auxiliary *will* ("If a person downloads child pornography, then the person will be punished for possession of child pornography"), then participants considered disablers – even for morally outraging offences. These studies show that the way a conditional is phrased is not ignored by participants. Instead, participants make inferences about the intended meaning of a particular phrasing. Usually, people consider their background knowledge about disablers and alternatives during reasoning. However, when the phrasing of a conditional suggests that background knowledge should not be used, then participants can refrain from using this knowledge. That is, the phrasing of a conditional can affect people's consideration of disablers and alternatives. The aim of this paper is therefore to continue investigating the importance of phrasing and pragmatics in reasoning. The general question is whether the specificity of the premises can affect reasoning. However, the more specific question we investigate here is whether the specificity of the person or object in the premises (henceforth: the term) can also affect people's consideration of background knowledge and acceptance of inferences.

In the beginning of our paper, we phrased the very same inference in two different ways: in the first example, the term is *unspecific* ("If a person studies hard, then the person will do well on the test"), but in the second example the term is *specific* ("If Anna studies hard, then Anna will do well on the test"). Both inferences have the same content (studying hard $\rightarrow$ doing well on the test) and only differ in how specific the term is. However, we expect these terms to activate a reasoner's background knowledge to a different extent. In the first example, the term is unspecific: it does not refer to a specific person or object, but refers only to a general person. The conditional does therefore make a *general* claim about the relationship between the antecedent and the consequent. Consequently, such an unspecific term addresses directly the reasoners' general knowledge about the relation between p and q. Therefore, we expect people to use their general knowledge about disablers and alternatives, which should lead to lower acceptance ratings of valid and invalid inferences.

In the second example, instead, the term is specific: it is a person named Anna, which the reasoner does not know. From a syntactic point of view, such a specific name has no special function: it could be seen as a simple placeholder for any other randomly picked person. That is, such as "Anna" there could be any other name. So theoretically, reasoners could also apply all their knowledge about disablers and alternatives. However, from a pragmatic point of view, things change. According to Grice (1975) and Sperber and Wilson (1995), utterances have to be as informative as necessary and also relevant for the listener. That is, if somebody mentions the name of a person, that specification should have an informative value. So, if in our experiments, a specific name or object is mentioned, reasoners should infer that there is a reason for mentioning that specific term in that particular premise. Maybe, the conditional relationship is true for this specific term? After all, reasoners do not know the specific person or object and thus cannot be sure whether they can apply their background knowledge about disablers and alternatives with certainty. In other words, reasoners should be less confident about the extent to which their existing background knowledge about disablers and alternatives also applies to specific terms. Thus, they should not consider this background knowledge to the same extent as for unspecific terms. We predict that this should result in higher acceptance ratings of valid and invalid conditional inferences with specific than unspecific terms.

We now present three experiments. In Experiment 1, we used a between subjects design and constructed inferences with either unspecific terms (e.g., “If a person…”, “If a plant…”) or specific terms (e.g., “If Anna…”, “If the Renum-plant…”). In Experiment 2, we used a within subjects design and varied the specificity of the premises orthogonally. That is, we also constructed problems where the conditional and the categorical had terms of a different specificity (unspecific conditional and specific categorical, or vice versa). Finally, the third experiment was similar to Experiment 2, but with a more representative sample and another randomisation technique. The paper ends with a discussion on the importance of phrasing, pragmatics and background in reasoning.

Experiment 1

Methods

Participants

One hundred and eighty-nine participants took part in the experiment. They were recruited online via the university database, consisted of undergraduate and graduate students and were $M = 21.41$ years old (SD = 3.32). From these 189 participants, we had to exclude 30 participants who indicated at the end of the experiment to not have worked concentrated on the task or to have already participated in a similar study in the past. We also excluded 16 participants who reported to have knowledge on formal logic. Our final sample consisted thus of 143 persons (115 females). Seventy-four of them were randomly assigned to the condition with unspecific terms and 69 of them were randomly assigned to the condition with specific terms (see the Materials and Design section). All participants had the possibility to take part in a lottery for two gift cards.

Materials and design

For our materials, we first conducted a preliminary study. In this preliminary study, we took 18 conditionals from the existing literature (from Cummins, 1995; De Neys, Schaeken, & d’Ydewalle, 2002; Verschueren et al., 2005) and asked participants to generate in 90 seconds as many disablers and alternatives as they could (cf. Cummins, 1995). All conditionals in the preliminary study were phrased with unspecific terms. From these 18 conditionals, we selected 8 conditionals with relatively high amounts of disablers and alternatives that could be phrased with terms of different specificity. In the experiment, we then phrased these conditionals with either *unspecific* terms (e.g., “If a person makes a diet, then…”, “If a plant is fertilised, then…”), or *specific*

Table 1. Conditionals used in Experiment 1 (translated from German language).

Conditional	Example
Unspecific	If a person studies hard, then the person will do well in the test.
	If a person drinks much cola, then the person gets fat.
	If a person sits in the draught, then the person will catch a cold.
	If a plant is fertilised, then the plant grows quickly.
	If a person reads without glasses, then the person gets a headache.
	If a person is on a diet, then the person loses weight.
	If a street is slippery, then there will be accidents on that street.
	If a person turns on the air conditioner, then the person feels cold.
Specific	If Sarah studies hard, then Sarah will do well in the test.
	If Thomas drinks much cola, then Thomas gets fat.
	If Claudia sits in the draught, then Claudia will catch a cold.
	If the Renumplant is fertilised, then the Renumplant grows quickly.
	If Florian reads without glasses, then Florian gets a headache.
	If Laura is on a diet, then Laura loses weight.
	If the Mantustreet is slippery, then there will be accidents on the Mantustreet.
	If Stefan turns on the air conditioner, then Stefan feels cold.

Note: Each conditional was presented four times, once per inference type (MP, MT, AC, DA); in the specific condition, each time with a different name.

terms (e.g., "If Laura makes a diet, then…", "If the Renumplant is fertilised, then …"). All eight conditionals were presented four times, as MP, MT, AC and DA inferences (in the specific condition, each time with a different name). Overall, we had thus 32 problems in each specificity condition. An overview of the selected conditionals can be found in Table 1.

The experiment followed a 2 (specificity: unspecific vs. specific) × 4 (inference: MP vs. MT vs. AC vs. DA) mixed design. The specificity was varied between subjects and the kind of inference within subjects. Each problem consisted thus of (i) a conditional (either unspecific or specific), (ii) a corresponding categorical (unspecific or specific, respectively) and (iii) a conclusion presented together with a 7-point Likert scale for the participants' acceptance ratings. As in previous studies (e.g., Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003a; 2003b), we asked participants to indicate how sure they were that the conclusion could be drawn (1 = *very unsure*, 7 = *very sure*). The specificity of the term in the conditional was the same as in the categorical and conclusion. For example:

If a person is on a diet, then the person loses weight.

A person is on a diet.

The person loses weight.

How sure are you that this conclusion can be drawn?

Very unsure 1 – 2 – 3 – 4 – 5 – 6 – 7 *Very sure*

Procedure

The experiment was programmed and conducted online via SoSci Survey (www.soscisurvey.de; Leiner, 2014). All premises of the inference tasks were presented at once on the screen, together with a 7-point Likert scale for acceptance ratings ranging from 1 = *Very unsure* to 7 = *Very sure*. Premises were presented in black font and the conclusion in red font. Participants were instructed that no right or wrong answers exist and that they should answer according to what they think applies more, as they do in everyday situations (highlighted in bold). We also added two further sentences to the instructions telling participants that it is possible that they feel that some problems are shown repeatedly or are very similar, but that they should nonetheless try to answer each problem independently from each other. The 32 inference problems were presented in a random order, preceded by a practice trial consisting on one problem (one MP inference with an unspecific term, not used later on during the experiment).

Results

The participants' acceptance ratings were analysed with a 2 (specificity: unspecific vs. specific) × 4 (inference: MP vs. MT vs. AC vs. DA) mixed Analysis of Variance (ANOVA). *Very unsure* was coded with one point and *very sure* was coded with seven points. Descriptive data can be found in Figure 1.

The ANOVA revealed a main effect of inference, $F(2.39, 337.26) = 102.69$, $p < .001$, $\eta_p^2 = .421$. Overall, MP ($M = 5.25$; SD = 1.16) received higher acceptance ratings than MT ($M = 3.88$; SD = 1.31), $t(142) = 13.65$, $p < .001$, $d = 1.103$,[1] and also higher ratings than AC ($M = 3.76$; SD = 1.20), $t(142) = 12.32$, $p < .001$, $d = 1.258$ and DA inferences ($M = 3.62$; SD = 1.25), $t(142) = 13.72$, $p < .001$, $d = 1.347$. MT, AC and DA inferences did not differ significantly from each other (all $p > .018$; Bonferroni adjusted alpha: 0.0083). In addition, the ANOVA also revealed a main effect of specificity, $F(1, 141) = 14.21$, $p < .001$, $\eta_p^2 = .091$. As expected, acceptance ratings were higher for specific terms ($M = 4.43$; SD = 0.98) than for unspecific terms ($M = 3.85$; SD = 0.87). The interaction between inference and specificity was not significant, $F(2.39, 337.26) = 0.01$, $p = .998$, $\eta_p^2 < .001$.

Discussion

The aim of Experiment 1 was to show that people's acceptance of conclusions is affected by the specificity of the term. When inferences contain specific terms, such as the name of a specific person, we expected pragmatics to make it clear to people that this specific person is not a simple placeholder but a person they do not know and that is worth being mentioned. This should in turn make people uncertain whether they can apply their

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

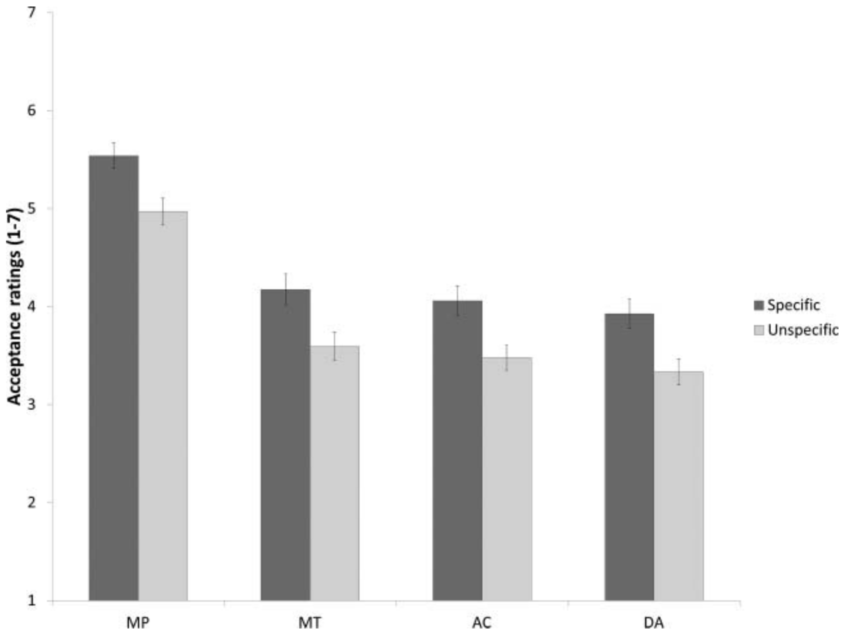

Figure 1. Acceptance ratings (1–7) for MP, MT, AC and DA inferences of different specificity (unspecific and specific) in Experiment 1. Error bars show standard errors.

knowledge about disablers and alternatives, which they had applied otherwise for unspecific terms. As expected, acceptance ratings were higher in the specific than unspecific condition, with a mean difference of 0.58 between both conditions. At first sight, this mean difference might seem rather small. However, further computations of standardised mean differences show that this difference is reliable ($d = 0.629$). In addition, the effect is highly significant and consistent for all kinds of inferences. This agrees with our assumption that the specificity of the term affects the consideration of one's background knowledge: When an inference contains a specific term, people do not know whether their background knowledge about disablers and alternatives applies. Thus, they refrain from considering this background knowledge and therefore accept more inferences.

This finding is important for cognitive psychology because the role of knowledge for reasoning has been mostly studied by varying the content of a conditional (e.g., familiar vs. unfamiliar content). However, we show that even for content for which we generally have knowledge and are familiar with, we can inhibit people's consideration of disablers and alternatives by phrasing the conditional with specific terms. This goes beyond a simple lack-of-knowledge effect and highlights the importance of pragmatics in reasoning. The knowledge people have about a certain domain of discourse is applied during reasoning only to the extent that reasoners think it is pragmatically acceptable to do so. When the phrasing of a conditional suggests that existing knowledge about disablers and alternatives is not applicable – for instance, when the specific term of a conditional is unknown – people seem to be less certain whether to apply this knowledge and show a higher acceptance of conclusions.

Experiment 2

In classical conditional inferences (e.g., "If *p* then *q*; *p*; therefore *q*"), the term in the conditional is expected to be the same as the one in the categorical. However, we can also create inferences where the term in the conditional is unspecific, but the one in the categorical is specific (henceforth: unspecific-specific problems), such as: "If a person makes a diet, then the person loses weight; Claudia is on a diet; Claudia loses weight". In our daily lives, we often make such inferences; for instance, when we apply general rules to specific observations, such as when we are trying to apply our knowledge about diets and healthy nutrition to individuals or also in domains which we expect to be highly deductive such as law – when general rules from penal code are applied to specific offenders (e.g., Gazzo Castañeda & Knauff, 2016a, 2016b). Making such unspecific–specific inferences seems intuitive because we know that the specific term in the categorical is actually an instance of the superordinate unspecific term in the conditional. In other words, saying e.g., "Claudia" or "a person" in the categorical statement should in principle not affect inferences, because both are "persons". But is this what actually happens when people are asked to make inferences? And what happens if we ask someone to make inferences from a conditional with a specific term to a categorical with an unspecific term (henceforth: specific–unspecific problems), such as: "If Tina is on a diet, then the Tina loses weight; A person makes a diet; A person loses weight"? From a classical perspective, this resembles more a case of induction and one can argue that these kinds of problems are qualitatively different from the other problems we used so far. However, nowadays the boundaries between deduction and induction in cognitive psychology are blurred (Evans, 2012) and there are many instances in our daily lives where we do such kind of inferences. For instance, when we observe that a specific person slips at the entrance of a slippery front yard, then we might conclude that other persons might also slip there. We know that from a perspective of probabilistic validity (p-validity), deductive and inductive inferences can still be distinguished (see Over, 2016). However, we never instructed our participants to decide about the validity of a conclusion, but only to reason as in daily lives. Therefore, we think that such problems are still worth investigating – even though we know that results on these problems should be interpreted with caution due to the invalidity of these inferences.

The aim of Experiment 2 was to reproduce the results of Experiment 1 (higher acceptance ratings for inferences with specific than unspecific terms) and to extend these results to tasks where participants draw inferences from unspecific terms to specific terms and vice versa. For the unspecific–specific inferences, we predict the participants to be unsure whether unspecific conditionals (e.g., "If a person…, then…") are also applicable to specific categoricals ("Anna…"), although this would be deductively valid. We therefore expect lower acceptance ratings in unspecific–specific conditions than in unspecific–unspecific conditions. For the specific–unspecific inferences, we predict the participants to not know if specific conditionals (e.g., "If Anna…, then…") are also applicable to unspecific categoricals ("A person…"). If the conditional applied to other persons, then according to Grice's (1975) maxim of quantity, the conditional should have been phrased with "a person" rather than with a specific name. Therefore, the acceptance ratings should be even lower than for unspecific–specific inferences.

Methods

Participants

We tested 24 students of the University of Giessen, but 3 had to be removed from the data-set because they reported to have visited courses on formal logic. The remaining 21 participants (13 females) were $M = 22.6$ years old on average (SD = 4.0).

Materials and design

For our materials, we took the same conditionals from Experiment 1. However, different to Experiment 1, this time, we decided to manipulate the specificity of the term in a within subject design. With these 8 conditionals, we therefore constructed 32 inference tasks with terms of different specificity. Eight inference tasks were *unspecific–unspecific* (unspecific term in the conditional and in the categorical), eight were *unspecific–specific* (specific term in the conditional and a specific term in the categorical), eight were *specific–unspecific* (specific term in the conditional and an unspecific term in the categorical) and eight were *specific–specific* (specific term in the conditional and in the categorical). We used the same eight conditionals for each of the four specificity conditions, but each time in a different inference form (to reduce the amount of items per participant). Which conditional was presented as MP, MT, AC or DA was decided randomly. As a result, each conditional was presented four times (once per specificity condition) and each inference eight times (twice per specificity condition). As in Experiment 1, the participants were asked to provide acceptance ratings by indicating how sure they were that the conclusion can be drawn (1 = *very unsure*, 7 = *very sure;* the order of the extremes was counterbalanced). Here is an example for an unspecific–specific problem:

If a person is on a diet, then the person loses weight.

Claudia is on a diet.

Claudia loses weight.

How sure are you that this conclusion can be drawn?

Very unsure 1 – 2 – 3 – 4 – 5 – 6 – 7 *Very sure*

The experiment followed a 4 (specificity: unspecific–unspecific vs. unspecific–specific vs. specific–unspecific vs. specific–specific) × 4 (inference: MP vs. MT vs. AC vs. DA) within subjects design.

Procedure

The experiment was programmed with Superlab 5.0 by Cedrus Cooperation. The participants were tested individually. During the instructions, the participants were told that they will be confronted with problems consisting of a rule, a fact and a conclusion, and that their task is to indicate how sure they were that this conclusion can be drawn. The participants were instructed that no right or wrong answers exist and that they had to answer as they do in everyday situations and according to what they think applies more (highlighted in bold and the word "everyday" in red). The conditional, categorical and conclusion were presented on separate screens. Participants could switch to the next screen by pressing the space bar. The conclusion was written in red font and appeared together with the 7-point Likert scale with which participants had to provide their acceptance ratings. The premises were written in black font. Between each problem, a fixation cross appeared. The 32 inference problems were presented in a random order, preceded by a short practice trial consisting of four problems, one for each inference form (MP, MT, AC, DA).

Results

Acceptance ratings were entered in a one-factorial repeated measures ANOVA with four levels (specificity: unspecific–unspecific vs. unspecific–specific vs. specific–unspecific vs. specific–specific). As in Experiment 1, *very unsure* was coded with one point and *very sure* was coded with seven points. We did not consider the factor "inference" in our analysis, because our randomisation technique did not allow for reliable comparisons in this respect. Since we used the same eight conditionals for each of the four specificity conditions, but each time in a different inference form (MP, MT, AC, DA), specificity-related comparisons within single inference levels are difficult (e.g., the content of MP–unspecific–unspecific inferences was not the same as in MP–

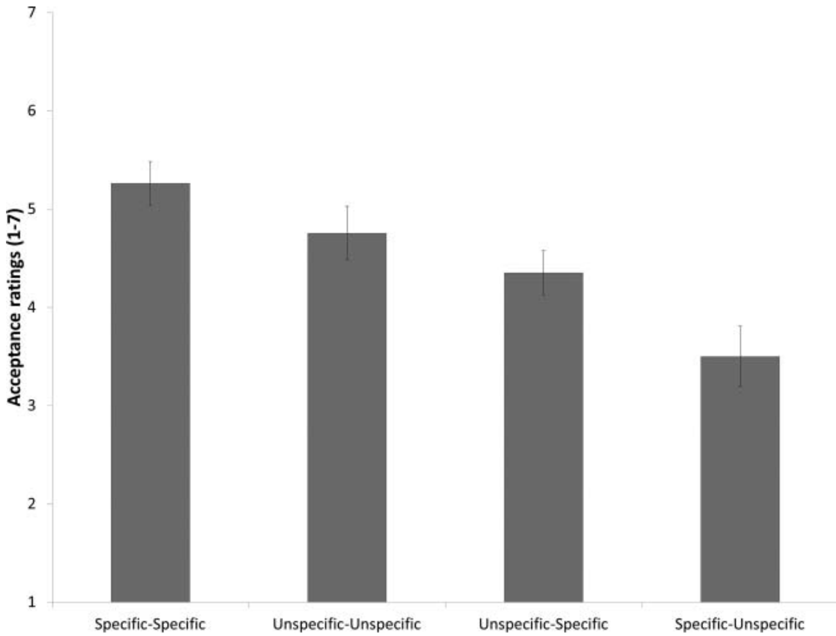

Figure 2. Acceptance ratings (1–7) for inferences of different specificity (specific–specific, unspecific–unspecific, unspecific–specific, specific–unspecific) in Experiment 2. Error bars show standard errors.

specific–specific inferences). That is, we *can* compare the overall acceptance ratings for the different specificity conditions (since they contain the same conditionals), but cannot do this on the inference level. Descriptive data can be found in Figure 2.

The ANOVA revealed a main effect of specificity, $F(1.89, 37.73) = 15.37$, $p < .001$, $\eta_p^2 = .435$. Descriptively, specific–specific problems ($M = 5.26$; SD = 1.03) received higher acceptance ratings than unspecific–unspecific problems ($M = 4.76$; SD = 1.25), $t(20) = 2.19$, $p = .020$, $d = 0.44$ (one-tailed), unspecific–unspecific problems received higher acceptance ratings than unspecific–specific problems ($M = 4.35$; SD = 1.04), $t(20) = 1.85$, $p = .040$, $d = 0.35$ (one-tailed) and unspecific–specific problems received higher acceptance ratings than specific–unspecific problems ($M = 3.5$; SD = 1.41), $t(20) = 4.03$, $p < .001$, $d = 0.65$ (one-tailed). Although not all pairwise comparisons reached the Bonferroni adjusted alpha level of 0.0167, this predicted linear trend was highly significant, $F(1, 20) = 23.93$, $p < .001$, $\eta_p^2 = .545$.

Discussion

Experiment 2 replicates the main findings of Experiment 1 and extends these findings to inferences where the term in the conditional and in the categorical is not the same. On the one hand, as in Experiment 1, we showed that conditionals with specific terms (here: specific–specific problems) receive higher acceptance ratings than conditionals with unspecific terms (here: unspecific–unspecific problems). On the other hand, we also found that acceptance ratings also depend on the relationship between the term in the conditional and the term in the categorical. When the specificity of the term in the conditional is not the same as the one in the categorical, acceptance ratings are lower; especially when a specific conditional is followed by an unspecific categorical. These results are in accordance with our predictions – even though the low acceptance rating for the specific–unspecific cases also might have been affected by the invalidity of the specific–unspecific inferences (see the Introduction of Experiment 2).

Overall, our results show that when reasoners have to decide about the acceptability of a conclusion, they usually take into account their background knowledge about the content of the conditional. However, this consideration of background knowledge can be modulated by pragmatic factors: when an inference is phrased with specific terms, this application of background knowledge is inhibited because reasoners do not know whether their background knowledge also applies for this specific case. Similar pragmatics are relevant when the specificity in the conditional and in the categorical mismatches. Reasoners know that background knowledge about one conditional relationship does not apply necessarily to other instances: disablers and alternatives applicable for unspecific conditionals do not have to apply to specific terms, and conditional relationships about specific persons or objects do not have to apply to other unspecified terms. That is, our findings show that conditional inference tasks are interpreted pragmatically as a set of assertions that have a reason to be asserted in a specific way. It is not simply the case that unspecific terms activate knowledge that is used no matter which additional premises follow. Instead, people take into account who is addressed in the conditional and who is addressed in the categorical and make judgements on whether the knowledge activated by the conditional is applicable to the categorical.

Experiment 3

In Experiment 2, we extended our findings from Experiment 1 to cases where the specificity of the conditional and the categorical is not the same. However, due to our randomisation technique, we were only able to make claims about the overall acceptance ratings for different specificity levels, but we were not able to investigate our specificity manipulation within different inference forms. In Experiment 3, we thus changed our randomisation technique by assigning randomly two conditionals to each inference form (MP, MT, AC or DA) and presenting these conditionals in all four specificity conditions (unspecific–unspecific vs. unspecific–specific vs. specific–unspecific vs.

specific–specific). This permitted us to test whether the findings from Experiment 2 can be replicated with this new randomisation technique and within the single inference forms.

Methods

Participants

One hundred and three participants took part in the experiment. They were recruited online via university databases and social networks and were M = 26.84 years old (SD = 6.55). Sixty-three per cent of the participants were graduate and undergraduate students, 30% were employees and 7% others. From these 103 participants, we had to exclude 17 participants who indicated at the end of the experiment to not have worked concentrated on the task. We also excluded 17 participants who reported to have knowledge in formal logic. Our final sample constituted thus of 69 persons (54 females). All participants had the possibility to take part in a lottery for two gift cards.

Materials and design

We used the same eight conditionals as in the previous experiments. This time, however, we distributed these eight conditionals randomly to the different inference forms: two conditionals were used for MP inferences, two for MT inferences, two for AC inferences and two for DA inferences. These two conditionals for each inference form were presented four times, once in each specificity condition, embedded in: *unspecific–unspecific*, *unspecific–specific*, *specific–unspecific* and *specific–specific* inference tasks. Overall, we thus had 32 problems. We know that this randomisation makes comparisons *between* the absolute acceptance rates for MP, MT, AC and DA inferences difficult (because different conditionals are used for each inference form). However, it nonetheless allows us to make specificity-related comparisons *within* MP, MT, AC or DA inferences. We can test how the acceptance of a conditional in one inference form changes depending on the specificity of the term (e.g., the unspecific diet-conditional in MP vs. the specific diet-conditional in MP). The experiment thus followed a 4 (specificity: unspecific–unspecific vs. unspecific–specific vs. specific–unspecific vs. specific–specific) $\times$ 4 (inference: MP vs. MT vs. AC vs. DA) within subjects design.

Procedure

The experiment was programmed and conducted online via SoSci Survey (www.soscisurvey.de; Leiner, 2014). As in Experiment 1, all premises of the inference tasks were presented at once on the screen, together with a 7-point Likert scale for acceptance ratings ranging from 1 = *Very unsure* to 7 = *Very sure*. Premises were presented in black font and the conclusion in red font. Again, the participants were instructed that no right or wrong answers exist and that they should answer according to what they think applies more, as they do in everyday situations (highlighted in bold). The participants were also told that it is possible that they feel that some problems are shown repeatedly or are very similar, but that they should nonetheless try to answer each problem independently from each other. All 32 problems were presented randomly and were preceded by one practice problem (one MP inference with an unspecific term, not used later on during the experiment).

Results

Acceptance ratings were analysed within a 4 (specificity: unspecific–unspecific vs. unspecific–specific vs. specific–unspecific vs. specific–specific) $\times$ 4 (inference: MP vs. MT vs. AC vs. DA) ANOVA for repeated measures. Again, *very unsure* was coded with one point and *Very sure* was coded with seven points. Descriptive data can be found in Figure 3.

As in the previous experiments, the ANOVA revealed a main effect of specificity, $F(1.93,\ 131.20) = 51.72$, $p < .001$, $\eta_p^2 = .432$. Overall, specific–specific problems ($M = 4.78$; SD = 1.04) received higher acceptance ratings than unspecific–unspecific problems ($M = 4.22$; SD = 1.03), $t(68) = 5.71$, $p < .001$, $d = 0.544$, unspecific–unspecific problems received higher acceptance ratings

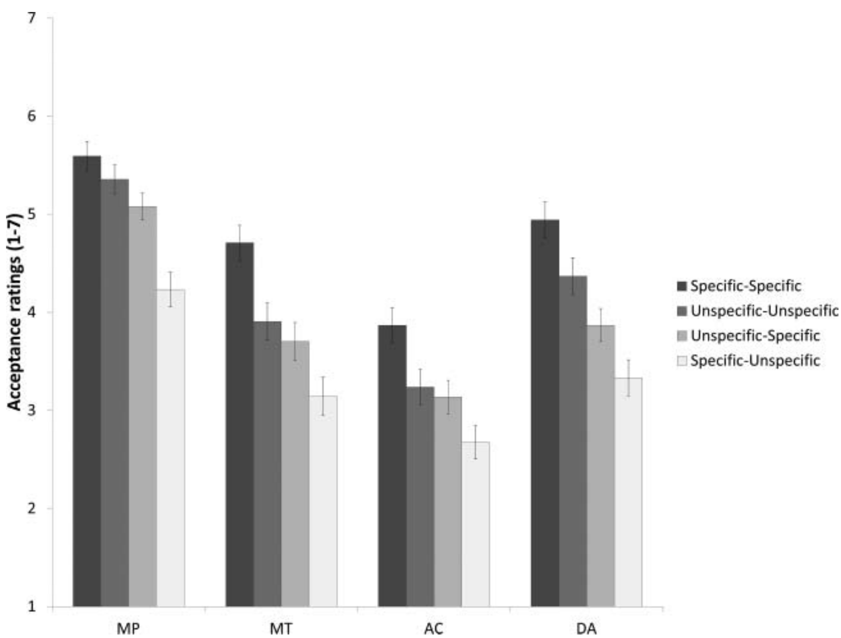

Figure 3. Acceptance ratings (1–7) for MP, MT, AC and DA inferences of different specificity (specific–specific, unspecific–unspecific, unspecific–specific, specific–unspecific) in Experiment 3. Error bars show standard errors.

than unspecific–specific problems (M = 3.95; SD = 1.01), t(68) = 4.14, p < .001, d = 0.265 and unspecific–specific problems received higher acceptance ratings than specific–unspecific problems (M = 3.35; SD = 1.21), t(68) = 4.95, p < .001, d = 0.533 (Bonferroni adjusted alpha: 0.0167). This linear trend was highly significant, F(1, 68) = 82.64, p < .001, η_p^2 = .549. In addition, the ANOVA revealed a main effect of inference, F(3, 204) = 53.48, p < .001, η_p^2 = .440, and an interaction between inference and specificity, F(7.34, 498.77) = 2.33, p = .022, η_p^2 = .033. The main effect of inference shows that overall MP inferences (M = 5.07; SD = 0.93) received higher acceptance ratings than MT inferences (M = 3.87; SD = 1.23), t(68) = 8.40, p < .001, d = 1.085, AC inferences (M = 3.23; SD = 1.26), t(68) = 10.76, p < .001, d = 1.641 and DA inferences (M = 4.13; SD = 1.21), t(68) = 7.30, p < .001, d = 0.851. Also, MT and AC inferences, t(68) = 4.33, p < .001, d = 0.510, and AC and DA inferences, t(68) = 6.00, p < .001, d = 0.725, differed significantly from each other, but MT and DA inferences did not, t(68) = 1.85, p = .069, d = 0.216 (Bonferroni adjusted alpha: 0.0083). However, due to the fact that we had different contents for each inference, this main effect should be interpreted with caution. Nevertheless, the interaction shows that the difference between specific–specific and unspecific–unspecific problems for MP inferences was not as pronounced as for the other inferences, t(68) = 1.76, p = .041, d = 0.19 (one-tailed; for MT, AC and DA: p < .001; Bonferroni adjusted alpha: 0.0125).

Discussion

The results of Experiment 3 agree with our previous findings and show once more that the specificity of the term affects people's acceptance of conclusions. The participants were confronted with inference tasks that were virtually identical in logical form and content. They only differed in the specificity of the term. However, this was enough to affect the reasoners acceptance ratings. Inference tasks in which the conditional and the categorical had specific terms received higher acceptance ratings than inference tasks in which the conditional and the categorical had unspecific terms. We also obtained lower acceptance ratings when the specificity of the conditional differed from the specificity of the categorical. We were therefore able to replicate the main findings of Experiment 2 also on the inference level. However, in contrast to Experiment 2, now the effect of specificity (specific–specific vs. unspecific–unspecific) was not that pronounced for MP inferences. Yet, a closer inspection of the data shows that this was primarily due to one of the two contents we used for MP inferences. For our MP inferences, we used the conditionals "If a person makes a diet, then the person loses weight" and "If a street is slippery, then there will be accidents on that street". For the diet-conditional, we can actually find the predicted difference between specific–specific (M = 5.80) and unspecific–unspecific inferences (M = 5.43). Only for the street-conditional this effect was very small (M = 5.39; M = 5.28; respectively). One reason for this may be that we used fictitious street names in order to create specific street-conditionals. We could not use real street names because this could have conflicted with background knowledge. So, maybe our fictitious street names (e.g., Mantustreet) were difficult to encode and thus the participants may have ended up encoding them as some unspecific "street" and not a specific street they do not know. This would also explain why the mean acceptance rate of this street-conditional in the specific form is very similar to the diet-conditional in its unspecific form.

General discussion

The aim of this paper was to show that the acceptance of conditional inferences depends on the specificity of the term. Across three experiments, we found that inferences with specific terms receive higher acceptance ratings than inferences with unspecific terms. In addition, in Experiments 2 and 3, we also showed that acceptance ratings are lower when the specificity of the conditional and categorical is not the same. We explain our findings by an interaction between pragmatics and background knowledge. When the term is unspecific, reasoners can apply all their background knowledge about disablers and alternatives, which consequently lowers acceptance ratings. However, when the term is specific (e.g., by having a specific name), pragmatics make it clear to people that this specific term is not a simple placeholder but a specific person or object they do not know. Consequently, reasoners do not know whether all the disablers and alternatives they know also apply to this specific case. Thus, they cannot apply their background knowledge about disablers and alternatives with certainty and this leads to higher acceptance ratings.

Our findings demonstrate how the phrasing of conditionals can affect inferences. This is an important message for all reasoning researchers. A review of the conditionals used in the literature shows that across studies – and even within studies – there is a lot of variation regarding the specificity of terms. Most of the studies on the role of disablers and alternatives employ conditionals with specific terms, such as "If Joyce eats candy often, then she will have cavities" or "If Louise turns on the air conditioning, then she will be cold" (e.g., Chan & Chua, 1994; Cummins, 1995; De Neys et al., 2002; Markovits, Forgues, & Brunet, 2012; Politzer & Bourmaud, 2002). Other conditionals are phrased in a more general way, such as "If a dog has fleas, then it will scratch itself" or "If water is heated to 100 °C, then it will boil" (e.g., Markovits et al., 2012; Verschueren et al., 2005; see also George, 1995). Yet others even use the reader, or an unknown third person as the term, e.g., "If you brush your teeth, then you will not get cavities" or "If she has an essay to write, then she goes to the library" (e.g., Byrne, 1989; Verschueren et al., 2005; see also

Dieussaert, Schaeken, & d'Ydewalle, 2002). Our findings show that such differences in the specificity of the term should not be ignored. Different levels of specificity can trigger different interpretations of the very same conditional. People seem to interpret conditionals with specific terms with a higher necessity and a higher sufficiency than conditionals with unspecific terms (cf. Thompson, 1994, 1995). Therefore, when planning experiments, researchers have to think carefully about how they phrase their materials. An arbitrary variation of the specificity of terms can lead to unwanted differences in specificity between conditions. And although the absolute difference between problems with specific and unspecific terms might look rather small (around 0.55 on a 7-point Likert scale), such an arbitrary variation of specificity can nonetheless lead to confounds. For instance, if one experimental condition presents more specific terms than another condition, then higher acceptance ratings of the first condition could be erroneously attributed to the experimental manipulation and not to the differing specificity of the terms.

Our findings are also relevant for the understanding of the role of pragmatics and background knowledge on reasoning. They show that pragmatics can moderate the consideration of background knowledge by either supporting or inhibiting the consideration of disablers and alternatives. This is more than a simple content effect. More than showing that people's everyday reasoning does not only depend on the structure of an inference (as formal logic suggests), our results show that pragmatics can moderate the consideration of background knowledge by either supporting or inhibiting the consideration of disablers and alternatives. In such a way, "content" is not only limited to the overall topic of a conditional, but is also related to the *specificity of the term*. We propose that this interaction between pragmatics and background knowledge happens in two phases. In the first phase, pragmatics make reasoners encode specific terms not as arbitrary terms, but as specific persons or objects they do not know but that seem to have specific properties which are relevant for the inference. Why else is the specification given? This thought probably causes the participants to interpret the conditional relationship as highly necessary and sufficient. Instead, if the term is unspecific, no specific name or labelling is given, then people interpret the conditional relationship as less necessary and sufficient. We call this the *pragmatic evaluation phase*. In a second step, the outcome of the pragmatic evaluation phase helps or inhibits the participants to consider their background knowledge about disablers and alternatives. We call this the *pragmatic application phase*. Of course, further studies are necessary to completely understand the relationship between pragmatics and background knowledge. One way to do so, for instance, could be the explicit presentation of disablers and alternatives. If the difference in acceptance ratings between specific and unspecific inferences is indeed due to less consideration of disablers and alternatives in the former than in the latter, then the explicit presentation of disablers and alternatives should have a higher impact on specific than on unspecific terms. Also, different kinds of "specificity" manipulations could be helpful for further testing the impact of pragmatics on reasoning. For instance, varying *how* the term of a conditional is addressed, such as in "If *a person* is on a diet, then the person loses weight" vs. "If *she* is on a diet, then the person loses weight". In this respect, it would also be interesting to test the impact of specific terms for which a reasoner does have knowledge. So far, we have assumed that specific terms are unknown and thus lower participant's willingness to apply their background knowledge about disablers and alternatives. This resulted from the fact that we were interested in cases where the specific term could be seen as a simple placeholder without affecting the potential disablers and alternatives for a given conditional. However, one can also imagine instances where people have concrete knowledge about specific terms; for instance, when the term is the reasoner himself (e.g., "If *you* study hard, then you will do well on the test"). In this case, reasoners would probably apply a very specific set of disablers and alternatives that can lead to either an enhanced or inhibited consideration of disablers and alternatives, depending on the individual experiences of each reasoner. It is a task for further studies to test the boundaries of the role of specificity on reasoning and to capture the conditions in which pragmatics lead to less consideration of disablers and alternatives in reasoning.

Another aspect that should be investigated in further studies is the phrasing of the dependent variable. In the present study, we measured acceptance ratings by asking the participants to indicate how sure they were that a conclusion can be drawn. We did that to make our findings comparable to other studies investigating the effect of disablers and alternatives on reasoning. For example, Cummins et al. (1991), De Neys et al. (2002, 2003a, 2003b), Cummins (1995) and others (e.g., Geiger & Oberauer, 2007; Markovits, Forgues, & Brunet, 2010; Vadeboncoeur & Markovits, 1999) also asked participants how sure or certain they were that a conclusion can be drawn. However, it would also be interesting to test other ways to measure conclusion acceptance. Although asking participants how sure they were that a conclusion can be drawn captures the acceptability of conclusions, it has also a metacognitive component which might have triggered the participants to think more deeply about the content – and the terms – of the inferences. In further studies, it would thus be interesting to replicate our findings either by asking for the acceptability of conclusions directly (Douven & Verbrugge, 2010) or for likelihoods or probabilities (e.g., Ohm & Thompson, 2004; Singmann & Klauer, 2011; Verschueren et al., 2005). Especially, the relation between specificity and probabilities should be investigated further. Many recent studies within the so-called new psychology of reasoning (Evans, 2012; Over, 2009) propose that conditionals are treated probabilistically (e.g., Evans & Over, 2004; Oaksford & Chater, 2007, 2013; Pfeifer, 2013). In such accounts, the subjective probability of the conditional, P(if p then q), is the conditional probability of q given p, $P(q|p)$. When evaluating, for instance, the acceptability of an MP conclusion, people first assume that p holds and then think how likely it is that q actually follows or not. The higher the subjective probability of P(p&q) is relative to P(p&not-q), the higher the conditional probability $P(q|p)$ is and the more readily an MP conclusion will be accepted (cf. Evans & Over, 2004; Over & Evans, 2003). But how can these probabilities capture the specificity of premises? And how are conditional probabilities computed for terms one does not know? We can imagine that specific terms elevate the conditional probabilities by also elevating the perceived necessity and sufficiency of p for q. But a germane question in this context is how the single probabilities of P(p&q) or P(p&not-q) are calculated cognitively, a question that is still unanswered. Background knowledge or the perceived frequency of occurrences cannot be solely responsible for these probabilities because for unknown specific terms no reliable background knowledge or frequency information exists.

We think that this is a new promising field of research. Long-time theories on conditional reasoning have tried to find overall explanations on how people reason with conditionals. However, considering the variety of situations in which we use conditionals in our daily lives, it is also necessary to look at the subtle differences in phrasing and pragmatics in order to fully understand how people interpret and reason with conditionals.

Acknowledgment

We thank Henry Markovits, Sangeet Khemlani and one anonymous reviewer for their helpful suggestions and comments on this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

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

References

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221–235). New York, NY: Russell Sage Foundation.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61–83.

Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition, 53*, 217–238.

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23*, 646–658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition, 19*, 274–282.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition, 30*, 908–920.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition, 31*, 581–595.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology, 15*, 161–176.

Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology, 49*, 181–195.

Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition, 117*, 302–318.

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning, 18*, 5–31.

Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford, UK: Oxford University Press.

Gazzo Castañeda, L. E., & Knauff, M. (2016a). Defeasible reasoning with legal conditionals. *Memory & Cognition, 44*, 499–517.

Gazzo Castañeda, L. E., & Knauff, M. (2016b). When will is not the same as should: The role of modals in reasoning with legal conditionals. *The Quarterly Journal of Experimental Psychology, 69*, 1480–1497.

Gazzo Castañeda, L. E., & Knauff, M. (2017). Quantifying disablers in reasoning with universal and existential rules. *Thinking & Reasoning*. doi: 10.1080/13546783.2017.1401000

Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition, 35*, 2060–2074.

George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86*, 93–111.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.

Hilton, D. J., Jaspars, J. M. F., & Clarke, D. D. (1990). Pragmatic conditional reasoning: Context and content effects on the interpretation of causal assertion. *Journal of Pragmatics, 14*, 791–812.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 636–678.

Leiner, D. J. (2014). SoSci Survey (Version 2.5.00-i) [Computer software]. Retrieved from https://www.soscisurvey.de

Markovits, H. (1986). Familiarity effects on conditional reasoning. *Journal of Educational Psychology, 78*, 492–494.

Markovits, H., Forgues, H. L., & Brunet, M.-L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition, 38*, 485–492.

Markovits , H., Forgues, H. L., , Brunet, M.-L. (2012). More evidence for a dual-process model of conditional reasoning. Memory & Cognition, *40*, 736–747.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning, 19*, 346–379.

Ohm, E., & Thompson, V. A. (2004). Everyday reasoning with inducements and advice. *Thinking & Reasoning, 10*, 241–272.

Over, D. (2016). The paradigm shift in the psychology of reasoning, In L. Macchi, M. Bagassi, & R. Viale (Eds.), Cognitive unconscious and human rationality (pp. 79–98). Cambridge: MIT Press.

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning, 15*, 431–438.

Over, D. E., & Evans, J. St. B. T. (2003). The probability of conditionals: The psychological evidence. *Mind & Language, 18*, 340–358.

Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking & Reasoning, 19*, 329–345.

Politzer, G., & Bourmaud, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology, 93*, 345–381.

Singmann, H., & Klauer, K. C. (2011). Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & Reasoning, 17*, 247–281.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford, UK: Blackwell Publishers Ltd.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition, 22*, 742–758.

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology, 49*, 1–58.

Vadeboncoeur, I., & Markovits, H. (1999). The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Thinking & Reasoning, 5*, 97–113.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning, 11*, 239–278.