The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

When will is not the same as should: The role of modals in reasoning with legal conditionals

Lupita Estefania Gazzo Castañeda & Markus Knauff

To cite this article: Lupita Estefania Gazzo Castañeda & Markus Knauff (2016) When will is not the same as should: The role of modals in reasoning with legal conditionals, The Quarterly Journal of Experimental Psychology, 69:8, 1480-1497, DOI: 10.1080/17470218.2015.1085067

To link to this article: http://dx.doi.org/10.1080/17470218.2015.1085067

Accepted author version posted online: 20 Aug 2015.
Published online: 27 Oct 2015.

Submit your article to this journal

Article views: 58

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=pqje20

Download by: [Universitaetsbibliothek Giessen] **Date:** 12 April 2016, At: 01:25

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
Vol. 69, No. 8, 1480–1497, http://dx.doi.org/10.1080/17470218.2015.1085067

Routledge
Taylor & Francis Group

When *will* is not the same as *should*: The role of modals in reasoning with legal conditionals

Lupita Estefania Gazzo Castañeda and Markus Knauff

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

(Received 2 March 2015; accepted 14 July 2015; first published online 27 October 2015)

Naïve reasoners reject logically valid conclusions from conditional rules if they can think of exceptions in which the antecedent is true, but the consequent is not. However, when reasoning with legal conditionals (e.g., "If a person kills another human, then this person should be punished for manslaughter") people hardly consider exceptions but evaluate conclusions depending on their own sense of justice. We show that participants' reluctance to consider exceptions in legal reasoning depends on the modal auxiliary used. In two experiments we phrased legal conditionals either with the modal "should" (i.e., " . . . then this person should be punished"), or with "will" (i.e., " . . . then this person will be punished") and presented them as modus ponens or modus tollens inferences. Participants had to decide whether the offender should or will be punished (modus ponens) or whether the offender indeed committed the offence (modus tollens). For modus ponens inferences phrased with "should" we replicate previous findings showing that participants select conclusions on the basis of their own sense of justice (Experiments 1 and 2). Yet, when the legal conditional is phrased with the modal "will" this effect is attenuated (Experiments 1 and 2), and exceptions are considered (Experiment 1). The modal auxiliary did not affect modus tollens inferences.

Keywords: Conditionals; Reasoning; Modals; Legal reasoning.

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

If the sun shines (p), Jack puts sunscreen on his face. (q)

The sun shines. (p)

Does Jack put sunscreen on his face? (q?)

What would you infer? According to classical logic the correct answer is "yes". This inference is called modus ponens (MP) and states that if the antecedent p is true, then the consequent q necessarily follows. Only the syntactic structure of the conditional rule matters, irrespective of its content. However, when participants are confronted with such inference tasks, they often do not rely on classical logic (Bonnefon & Vautier, 2010; Evans, 2002, 2012; Oaksford & Chater, 2009). Naïve reasoners often activate their knowledge about the content of the conditional and consider information that is not explicitly given in the premises (e.g., De Neys, Schaeken, & d'Ydewalle 2003a, 2003b; Dieussaert, De Neys, & Schaeken, 2005; Evans & Over, 2004; Johnson-Laird & Byrne,

Correspondence should be addressed to Lupita Estefania Gazzo Castañeda, Experimental Psychology and Cognitive Science, Justus Liebig University Giessen, Otto-Behaghel-Str. 10F, 35394 Giessen, Germany. E-mail: Estefania.Gazzo@psychol.uni-giessen.de

We thank Jean-Francois Bonnefon for the inspiration on Experiment 2, and Jonathan Evans and two anonymous reviewers for their helpful comments and suggestions on earlier versions of this paper.

This research was supported by DFG grants [grant number KN 465/10-1], [grant number KN 465/10-2] to Markus Knauff within the DFG Priority Program "New Frameworks of Rationality" (SPP1516).

© 2016 The Experimental Psychology Society

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

2002; Oaksford & Chater, 1995, 2003, 2009). By doing that, they start to think of exceptions: situations where even though the antecedent *p* of the conditional is true, the consequent *q* is not (e.g., Byrne, 1989; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys et al. 2003a, 2003b; Dieussaert, De Neys, & Schaeken, 2005; Johnson-Laird & Byrne, 2002; Markowitz & Potvin, 2001; Oaksford & Chater, 2001). The more frequently they perceive exceptions to occur, the less likely they think it is that *q* follows from *p*, and hence the less they accept a conclusion from a conditional rule, although it is logically correct (Dieussaert, Schaeken, and d'Ydewalle, 2002; Geiger & Oberauer, 2007; Oaksford & Chater, 2001, 2003; Weidenfeld, Oberauer, & Hörnig, 2005; see also Over & Evans, 2003). Many researchers working on the "new paradigm" of cognitive psychology (Evans, 2012) actually propose that conditionals are understood probabilistically and that the probability of a conditional depends on the conditional probability of *q* given *p* (Baratgin, Over, & Politzer, 2014; Elqayam & Over, 2013; Evans & Over, 2004; Oaksford & Chater, 2010; see also Edgington, 1995).

In our initial example, naïve reasoners may thus refuse to conclude "yes" if they can think of, for example, Jack being inside his house or his bottle of sunscreen being empty.

Exceptions may also affect other kinds of inferences, like the modus tollens (MT) inference:

If the sun shines (*p*), Jack puts sunscreen on his face. (*q*)

Jack doesn't put sunscreen on his face (not-*q*)

Does the sun shine (*p*?)

The logically correct answer to MT is "no", since according to classical logic the antecedent of a conditional is sufficient but not necessary for the consequent to occur. However, as in the prior example, people may refuse to conclude "no" if they are aware that cases exist where Jack does not put sunscreen on his face although the sun is shining.

The effect of exceptions on reasoning is supported by a vast range of experiments (e.g., Byrne, 1989; Chan & Chua, 1994; Cummins, 1995; Cummins et al., 1991; DeNeys et al., 2003a, 2003b; Fernbach & Erb, 2013; Manktelow & Fairley, 2000; Pijnacker, Geurts, Lambalgen, Buitelaar, & Hagoort, 2010; Politzer & Bourmaud, 2002; see also Oaksford & Chater, 2001, 2003; Oaksford, Chater, Larkin, 2000). However, it seems not to apply to every kind of conditional. Contrary to what is found in most of the literature, Gazzo Castañeda and Knauff (2015) showed that people have difficulties in accepting exceptions when confronted with *legal conditionals* such as "If a person kills another human, then this person should be punished for manslaughter" and are asked to infer whether the offender should be punished. In such problems people typically refuse to accept legal exceptions to rules (e.g., self-defence or lack of criminal liability due to psychological disorders), in particular if the offence in the legal conditional is of high moral outrage. Why so?

One reason might be the importance of punishing offenders in our society. Research on social justice shows that people experience negative feelings of *moral outrage* when faced with offences, resulting in a desire of punishment (Alter, Kernochan, & Darley, 2007; Carlsmith & Darley, 2008; Carlsmith, Darley, & Robinson, 2002; Darley, 2009; Darley & Pitmann, 2003; also see E. Fehr & Gächter, 2002; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001, 2007; Keller, Oswald, Stucki, & Gollwitzer, 2010).

Another reason—which we explore here—might be that the modal auxiliary used in the legal conditional can also motivate participants to ignore the availability of exceptions. In the experiments of Gazzo Castañeda and Knauff (2015), legal conditionals were phrased with the modal "should", and not with "will" as it is the case in most of the literature on conditional reasoning (Kilpatrick, Manktelow, & Over, 2007). In legal theory, modals such as "should" or "ought" stand for the normative nature of legal rules (Bäcker, 2009, 2010). However, when presented in a MP inference, the modal "should" can have additional implications, as can be seen in the following example:

If a person kills another human, then this person should be punished for manslaughter.

A person killed another human.

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

Should this person be punished for manslaughter?

The "should" in the conclusion can be understood as asking for what *should* happen according to the deontic principles of this rule (see deontic possibilities; Johnson-Laird & Byrne, 2002; Quelhas & Byrne, 2003), or, one step further, according to the own deontic standards of the reasoner. Participants might ignore exceptions because they answer in line with what is correct according to their own sense of justice. Over, Manktelow, and Hadjichristidis (2004) already argued that for deontic conditionals, the acceptance of conditional rules does not depend on the perceived probability of *q* given *p*, as is the case for factual conditionals, but instead on the preference of the different outcomes such a rule can have. The more a reasoner prefers the outcome "*p* and *q*" (in our case: committing an offence and being punished for that) over "*p* and not *q*" (in our case: committing an offence and not being punished for that), the more the rule will be accepted. In the case of deontic legal conditionals, we expect the preference of "*p* and *q*" over "*p* and *not q*" to be correlated with moral outrage: The more morally outraged a reasoner is by the offence in the legal conditional, the more she or he will prefer the outcome "*p* and *q*" (i.e., offence and punishment) over "*p* and not *q*" (i.e., offence and no punishment), and the more she or he will conclude that the offender should be punished.

But what happens if instead of "should", the legal conditional is phrased with the modal "will"? Consider the following example:

If a person kills another human, then this person will be punished for manslaughter.

A person killed another human.

Will this person be punished for manslaughter?

Different from "should", the modal "will" in the conclusion suggests that the inference is asking about what happens in the real world—that is, what is factually the case (see factual possibilities, Johnson-Laird & Byrne, 2002; Quelhas & Byrne, 2003). Therefore, when thinking about what happens in real world, the own sense of justice should lose importance, and exceptions should gain importance. The more frequently participants perceive exceptions to occur for a given offence, the less they will think *q* happens given *p*, and the less the rule will be accepted. In this case, when asked whether an offender will be punished, a reasoner might conclude that the offender will not be punished even though the offence is severe and morally outrageous.

The fact that different modals can have different implications—either by their meaning per se or by the context in which they are uttered—is known from linguistics (e.g., Groefsema, 1995) and has also received some psychological support (e.g., Johnson-Laird, 1978). Bell and Johnson-Laird (1998), for instance, showed that depending on which modal is used ("can" vs. "must"), inferences are drawn differently fast: When asked about what *can* be the case, participants take longer to answer affirmatively than to answer negatively, but when asked about what *must* be the case it is the other way around. Further, Ferguson and Sanford (2008) showed that modals in counterfactual conditionals can affect interpretation of subsequent information. Finally, Elqayam, Thompson, Wilkinson, Evans, and Over (in press) showed that modals can be used when reasoning with utility conditionals and deciding whether an action should be taken or not. In fact, the difference between asking for deontic states or factual states is widely known in the psychological literature (e.g., Beller, 2008; Bucciarelli & Johnson-Laird, 2005). However, aside from the above-mentioned studies, most of the research about the difference between factual and deontic reasoning has been done with Wason's (1968) selection task where participants have to choose cards in order to falsify or violate a rule (Beller, 2008; Kilpatrick et al., 2007; Manktelow & Over, 1991; see Cheng & Holyoak, 1985; Cosmides, 1989).

The role of modals in the consideration of exceptions in inference tasks has thus not received much attention yet. To fill this gap, the aim of this study is to investigate the effect of modals on the consideration of exceptions in legal conditional reasoning. Because of the implications of asking about "should" or "will" we predict different answer patterns depending on how the legal conditional is phrased. If the legal conditional has the modal "should" (deontic legal conditionals), people should base their conclusions on their own sense of justice

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

and feelings of moral outrage, ignoring the existence of exceptions. This should reflect in the participants' acceptance of a rule, given their preference of "*p* and *q*" over "*p* and *not q*" (see Over et al., 2004). However, if the legal conditional has the modal "will" (factual legal conditionals), people should base their conclusions on their knowledge about what happens in the real world, leaving their own sense of justice behind and considering exceptions. This consideration of exceptions should be reflected in the perceived frequency of cases of "*p* but *not q*" (cf. Geiger & Oberauer, 2007).

We tested our hypotheses in two experiments. In the first experiment, participants were confronted with legal conditionals incorporated in MP and MT inferences. We varied the modal, the frequency of exceptions[1] and the participants' acceptance of the rule (i.e., the relative preference of "*p* and *q*" over "*p* and *not q*"). In the second experiment, a different set of participants was confronted with the same conditionals as those in the first experiment, but the offenders described in the legal conditionals were now replaced by close others (family members and friends), which people usually forgive more easily and do not want bad things to happen to (e.g., Buckley, Chapman, Sheehan, & Cunningham, 2012; R. Fehr, Gelfand, & Nag, 2010; McCullough et al., 1998). If one's own deontic preferences influence only deontic conditionals, but not factual conditionals, then putting close people as offenders should affect conclusions for deontic legal conditionals but not for factual ones.

EXPERIMENT 1

Method

Participants

A total of 42 participants took part in the experiment. We excluded two participants from the analysis because they afterwards reported to have specialized knowledge of logic or about the German penal code. The final sample consisted of 40 participants (20 female), with a mean age of 23.7 years ($SD = 3.0$). Half of the participants received deontic legal conditionals, the other half factual legal conditionals.

Material and design

The materials were selected through a large ($N = 298$) preliminary study via SoSci Survey (Leiner, 2014). In the first half of the preliminary study, we measured the participants' acceptance of legal rules. Based on the proposal of Over et al. (2004), participants were confronted with legal rules together with their outcomes "*p* and *q*" and "*p* and *not q*" (e.g., "A person kills another human and is punished for manslaughter" and "A person kills another human and is not punished for manslaughter"). In total we tested $N = 92$ legal rules but each participant received only 14–16 of them. The participants' task was to indicate their preference for each of these two possible outcomes on a scale from 1 (*completely in favour*) to 7 (*completely against*). We computed a participant's acceptance of a rule by dividing the ratings she or he gave for the outcome "*p* and *not q*" by the ratings for the outcome "*p* and *q*". The higher this quotient, the more a participant accepted a rule.

In the second half of the preliminary study, participants were asked to rate the frequencies of exceptions for each rule. For this they were confronted once more with the same rules and were asked to rate in how many of 100 cases they thought *p* (i.e., the offence) occurs, but without the following *q* (i.e., the punishment; e.g., "A person kills another human, but the person is not punished for manslaughter. In how many of 100 cases do you think this is the case?").

The offences we finally used for the conditionals in the actual experiment were selected depending on

[1]Strictly speaking, for deontic conditionals instances of *p* and ¬*q* are not exceptions, but violations (e.g., Beller, 2008, 2010). If somebody kills somebody else but is not punished, this can be considered a violation of the manslaughter rule, for instance when the lack of punishment is due to malpractice. However, for legal conditionals this is not always the case. Penal code includes several instances that actually permit cases of *p* and ¬*q*. For instance, if somebody is not punished for manslaughter because of self-defence, it is not a violation to the manslaughter rule, but an exception. Given this peculiarity of legal conditionals, and in order to facilitate readability, in this paper we use the word "exception" for both factual and deontic conditionals.

Table 1. *Legal conditionals used in Experiment 1, together with means (and standard deviations) for rule acceptance (RA) and frequency of exceptions (EX) from the preliminary studies*

Items	*RA (SD)*	*EX (SD)*
High RA, high EX (high-high)		
1. If a person downloads child pornography, then this person should/ will be punished for possession of child pornography.	6.3 (1.5)	56.6 (27.9)
2. If a person pollutes the soil and thereby harms animals, then this person should/ will be punished for soil pollution.	5.4 (2.2)	70.9 (24.9)
High RA, low EX (high-low)		
1. If a person abducts a human being in order to coerce a third person to commit an act, then this person should/ will be punished for hostage taking.	6.5 (1.3)	19.9 (17.4)
2. If a person kills another human, then this person should/ will be punished for manslaughter	5.8 (2.1)	22.9 (22.1)
Low RA, high EX (low-high)		
1. If a person downloads music from the internet without allowance, then this person should/ will be punished for breaching the copyright law.	1.9 (2.2)	69.9 (34.7)
2. If a person participates in an illegal game of chance, then this person should/ will be punished for illegal gambling.	2.1 (1.8)	62.2 (25.5)
Low RA, low EX (low-low)		
1. If a person kills another human because of the explicit and earnest request of the person killed, then this person should/ will be punished for homicide upon request.	2.0 (2.3)	14.8 (21.6)
2. If a shop-owner opens his/ her shop without allowance on a Sunday, then this person should/will be punished for breaching the Shop Closing Act.	1.9 (2.0)	26.5 (26.3)

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

the ratings obtained in both parts of the preliminary study. We selected eight offences: two with high acceptance rates and high frequency of exceptions, two with high acceptance and low frequency of exceptions, two with low acceptance and high frequency of exceptions, and two others with low acceptance and low frequency of exceptions. The assignment to each category was corroborated statistically. The list of the legal conditionals used in the experiment together with their ratings from the preliminary study can be found in Table 1.

For the experiment, each offence was phrased as a legal conditional and was presented once in an MP and once in an MT inference, creating a total of 16 problems. The conclusion was phrased as a question. Thus the inference problems consisted of (a) a legal conditional rule, (b) the fact *p* or *non q* (for MP and MT inferences, respectively), and (c) the question about the conclusion, asking whether *q should* or *will* follow (for MP inferences) or whether *p* is the case (for MT inferences). Half of the participants got the problems with the modal "should" and the other half with "will". See Table 2 for an illustration. Thus the experiment followed a 2 (modal: should vs. will) × 2 (rule acceptance: high vs. low) × 2 (frequency of exceptions: high vs. low) × 2 (inference: MP vs. MT) mixed design. The modal was varied as a between-subjects factor; all other factors were varied within individuals.

We also created a generation and an evaluation task. The generation task served as an additional measure for the availability of exceptions. We presented participants the same eight offences from the inference task and asked them to generate reasons of why somebody who committed those offences *should/will* not be punished ("A person kills another human, but this person should/will not be punished for that"). The modal used in the generation task was the same as the one that the participants had received in the inference task. Following previous studies (see Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003a, 2003b) participants had 1.5 min to write down as many reasons as they could think of. After 10 s of inactivity, the 1.5 min ended prematurely.

Table 2. *Structure of the problems used in Experiment 1 illustrated by the legal conditional of manslaughter*

Modal		*Modus ponens*		*Modus tollens*
Should	R	If a person kills another human, then this person should be punished for manslaughter.	R	If a person kills another human, then this person should be punished for manslaughter.
	F	A person kills another human.	F	A person should not be punished for manslaughter.
	C	Should this person be punished for manslaughter?	C	Did this person kill another human?
Will	R	If a person kills another human, then this person will be punished for manslaughter.	R	If a person kills another human, then this person will be punished for manslaughter.
	F	A person kills another human.	F	A person is not punished for manslaughter
	C	Will this person be punished for manslaughter?	C	Did this person kill another human?

Note: R = conditional rule; F = fact; C = conclusion.

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

In the evaluation task, participants were confronted once more with the same eight offences and were asked to indicate on a 7-point Likert scale how morally outraged they were by each (1 = *no moral outrage*, 7 = *high moral outrage*). Since feelings of moral outrage towards offences are an essential part of people's sense of justice (Carlsmith & Darley, 2008; Carlsmith et al., 2002; Darley, 2009; Darley & Pitmann, 2003), we expected the moral outrage ratings to correlate with the acceptance ratings from the preliminary study.

Procedure

Participants were tested individually on a computer. The experiment was presented with Cedrus Superlab © 4.5 and was introduced as an experiment on reasoning in law. The participants were told that they will be confronted with general rules that are embedded in specific situations and that their task is to decide whether this rule *should* (for participants in the deontic legal conditionals condition) or *will* (for participants in the factual legal conditional condition) be applied in the given situation. Apart from the modal used for describing the task, instructions were kept constant across conditions. After one practice trial consisting of two items with the legal conditional for tax aversion (once as MP, once as MT), participants were left alone in the experimental room. The two premises were presented on subsequent screens, and participants could move to the next screen by pressing the space bar. The conclusion was always phrased as a question and was written in red. After reading this question, participants had to answer on a 7-point Likert scale ranging from *yes, very certainly* to *no, very certainly not*. The polarity of the scale was reversed for half of the participants. Between each problem, participants had the opportunity to take a break. After the inference task, the two supplementary tasks were presented in random order. After the experimental session, participants were asked in an open interview about their knowledge about logic and law.

Results

Inference task

For the MP inferences "yes, very certainly" answers were scored with 0 points and "no, very certainly not" answers with 6 points. The ratings in-between were scored respectively with 1 to 5 points. For the MT inferences the opposite was done. "No, very certainly not" answers were scored with 0 points and "yes, very certainly" with 6 points. Again, the ratings in-between were scored respectively with 1 to 5 points. These scores were averaged separately for MP and MT inferences and indicate the degree of rejection of the logically valid conclusion. We call this the "rejection rating". The higher the rejection rating, the less the MP or MT inferences were accepted. The rejection ratings for MP and MT inferences can be found in Figure 1.

A 2 (modal: should vs. will) × 2 (rule acceptance: high vs. low) × 2 (frequency of exceptions: high vs. low) × 2 (inference: MP vs. MT) analysis of variance (ANOVA) on rejection ratings showed a main effect of frequency of exceptions,

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

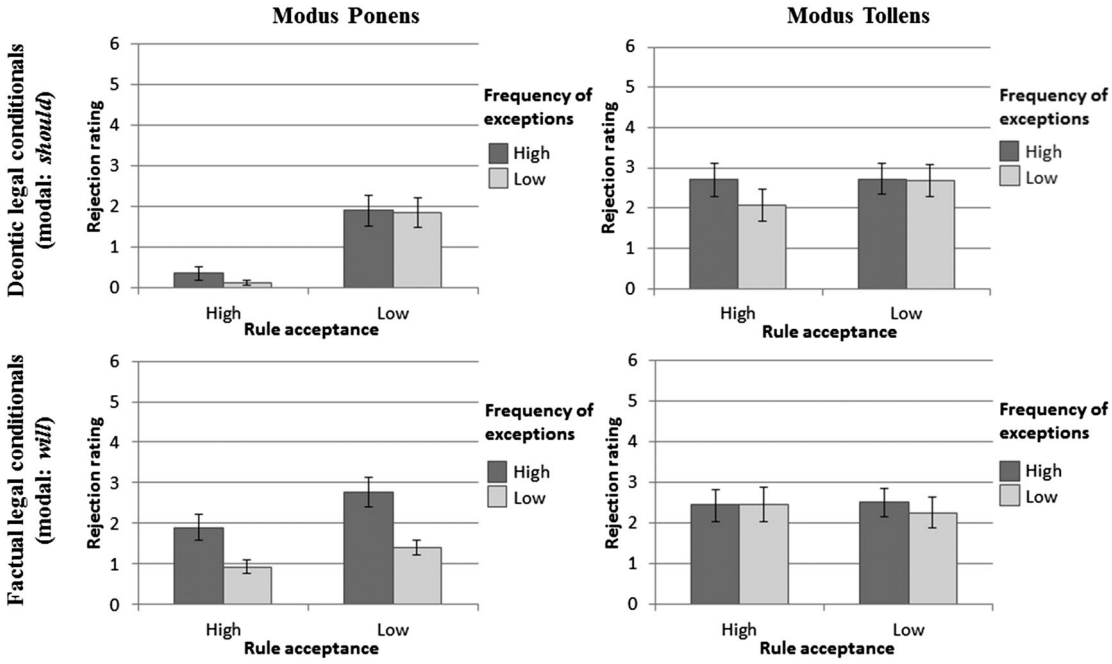

Figure 1. *Rejection ratings (0–6) for modus ponens and modus tollens inferences for deontic and factual legal conditionals in Experiment 1. Error bars represent standard errors.*

$F(1, 38) = 14.06$, $p = .001$, $\eta_p^2 = .270$, a main effect of rule acceptance, $F(1, 38) = 23.05$, $p < .001$, $\eta_p^2 = .378$, and a main effect of inference, $F(1, 38) = 21.69$, $p < .001$, $\eta_p^2 = .363$. However, the ANOVA also showed that the kind of inference (MP vs. MT) interacted significantly with rule acceptance, $F(1, 38) = 30.03$, $p < .001$, $\eta_p^2 = .441$, with the frequency of exceptions and the modal auxiliary, $F(1, 38) = 6.73$, $p = .013$, $\eta_p^2 = .15$, and marginally with the frequency of exceptions, $F(1, 38) = 3.08$, $p = .087$, $\eta_p^2 = .075$, and with the modal auxiliary, $F(1, 38) = 3.128$, $p = .085$, $\eta_p^2 = .076$. Therefore we decided to analyse the data from MP and MT inferences with two separate 2 (modal: should vs. will) × 2 (rule acceptance: high vs. low) × 2 (frequency of exceptions: high vs. low) ANOVAs. As can be seen in Figure 1, results for MP and MT indeed differed.

For MP inferences we found main effects for modal auxiliary, $F(1, 38) = 5.78$, $p = .021$, $\eta_p^2 = .132$, for rule acceptance, $F(1, 38) = 43.988$, $p < .001$, $\eta_p^2 = .537$, and for frequency of exceptions, $F(1, 38) = 20.021$, $p < .001$, $\eta_p^2 = .345$. These effects were explained in terms of the expected interactions between the modal auxiliary and rule acceptance, $F(1, 38) = 7.62$, $p = .009$, $\eta_p^2 = .167$, and between the modal auxiliary and the frequency of exceptions, $F(1, 38) = 12.51$, $p = .001$, $\eta_p^2 = .248$. On the one hand, the interaction between the modal auxiliary and frequency of exceptions shows that the frequency of exceptions had a strong effect on factual legal conditionals, but did not have any effect on deontic legal conditionals: Whereas rejection ratings were higher for factual conditionals with high frequency of exceptions ($M = 2.34$, $SD = 1.43$) than for those with low ($M = 1.16$, $SD = 0.60$), $t(19) = 4.91$, $p < .001$, $d = 0.81$,[2] rejection ratings for deontic conditionals with high ($M = 1.13$, $SD = 1.05$) and low frequency of exceptions ($M = 0.99$, $SD = 0.83$) did not differ, $t(19) = 0.49$, $p = .428$, $d = 0.14$ (Bonferroni adjusted alpha level, $\alpha = .025$). On

[2]Standardized mean differences (d) were computed as described by Borenstein (2009).

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

the other hand, the interaction between modal auxiliary and rule acceptance shows that the effect of rule acceptance on rejection ratings was higher for deontic legal conditionals than for factual legal conditionals: Although rejection ratings for conditionals with low rule acceptance were always higher than rejection ratings for conditionals with high rule acceptance, this effect was higher for deontic ($M = 1.88$, $SD = 1.48$ vs. $M = 0.24$; $SD = 0.43$; respectively), $t(19) = 5.58$, $p < .001$, $d = 1.24$, than for factual conditionals ($M = 2.09$, $SD = 1.13$ vs. $M = 1.41$, $SD = 0.95$, respectively), $t(19) = 3.58$, $p = .002$, $d = 0.64$ (Bonferroni adjusted alpha level, $\alpha = .025$). All other effects were not significant (all $F \leq 1.6$, $p \geq .21$).

The ANOVA for the MT inferences did not show any significant effects at all (all $F \leq 1.60$, $p \geq .214$).

Generation task

Two raters counted independently the amount of exceptions generated by participants and also the quality of each exception ($\rho = .98$ for the amount of exceptions in general, $\rho = .98$ for amount of factual, and $\rho = .92$ for the amount of deontic exceptions). Exceptions describing cases of malpractice (e.g., not being caught, not being accused, influences, etc.) were counted as factual exceptions. Exceptions describing cases were an offender should not be punished as a matter of principle (e.g., lack of criminal liability) were counted as deontic exceptions. We analysed the amount of exceptions generated within a 2 (modal: should vs. will) × 2 (rule acceptance: high vs. low) × 2 (frequency of exceptions: high vs. low) ANOVA. We only found a main effect for the modal auxiliary, $F(1, 38) = 4.55$, $p = .04$; $\eta_p^2 = .107$, and a main effect of the amount of exceptions, $F(1, 38) = 5.25$, $p = .028$; $\eta_p^2 = .121$. Participants in the factual condition generated more exceptions than participants in the deontic condition ($M = 2.53$, $SD = 1.28$, vs. $M = 1.83$, $SD = 0.69$, respectively). Despite not being a big difference descriptively, participants also generated more exceptions for offences classified as having a low frequency of exceptions than for offences classified as having high frequency of exceptions ($M = 2.29$, $SD = 1.16$, vs. $M = 2.06$, $SD = 1.08$, respectively). We also found that the quality of exceptions differed depending on whether the participant was confronted with deontic or factual legal conditionals: While participants in the factual condition generated in 53% of the cases factual exceptions, participants in the deontic condition did this only in 20% of the cases. Along the same lines, deontic exceptions were more frequent in the deontic condition than in the factual condition (77% vs. 43%, respectively).

Moral outrage

We correlated the moral outrage ratings given to offences in the inference task with the corresponding rule acceptance ratings for the same offences from the preliminary study. As expected, the mean moral outrage ratings correlated with the mean rule acceptance ratings. This was the case for both the moral outrage ratings given in the "should" condition, $r = .95$, $p < .001$, and the ones in the "will" condition, $r = .97$, $p < .001$. The more the participants accept a legal rule, the more moral outrage they feel when this rule is broken.

Discussion

Results show that the modal auxiliary used in legal conditionals affects the MP inferences drawn. If the conditional is phrased with the modal auxiliary "should", participants seem to rely on their own sense of justice and feelings of moral outrage: Whereas for conditionals about severe offences participants show very low rejection ratings (i.e., favouring that the offender should be punished), for conditionals about mild offences—which they do not accept—they show higher rejection ratings (i.e., favouring that the offender should not be punished). This replicates the findings of Gazzo Castañeda and Knauff (2015). However, the effect is attenuated for conditionals phrased with the modal "will". For factual conditionals also the frequency of exceptions matters. Whereas for conditionals with a low frequency of exceptions participants show small rejection ratings (i.e., favouring that the offender will be punished), for conditionals with a high frequency of exceptions they show

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

higher rejection ratings (i.e., favouring that the offender will not be punished). Since this was the case for both severe and mild offences, the modal "will" seems to make people think that factual, and not deontic, principles are asked for. This is corroborated by the generation task, where factual exceptions were much more present in the factual condition than in the deontic condition. Yet, an unexpected result was that the conclusions drawn from factual conditionals were influenced not only by the frequency of exceptions, but also by the participants' acceptance of the rule. One probable explanation is that participants still followed their own sense of justice to some extent. This is in accordance with the results of Gazzo Castañeda and Knauff (2015), where lay people were still affected by their feelings of moral outrage even although instructed to decide like a real judge.

Contrary to the results for MP inferences, those for MT were not so straightforward. We expected to find the same pattern of results as that for MP inferences, but this was not the case. None of the factors influenced the conclusions selected for MT inferences. This is not the first time MT produces unexpected results (e.g., De Neys et al., 2003a; Johnson-Laird & Byrne, 1991; Singman, Klauer, & Over, 2014). A common explanation is that the negation in the MT inference makes the task more difficult (De Neys et al., 2003a). In our case, this higher difficulty might have led to more logical errors, which might have covered the effects of modals, exceptions and rule acceptance. In fact, participants showed higher rejection ratings for MT inferences ($M = 2.48$, $SD = 1.45$) than for MP inferences ($M = 1.40$, $SD = 0.97$), $t(39) = 4.54$, $p < .001$, $d = 0.86$.

An alternative explanation is that the higher rejection ratings for MT inferences are because participants were in some way led to assume that p was the case although q was not. According to Sperber and Wilson's (1995) principle of relevance, people make inferences about the messages they get assuming that the information is relevant to them. In this way, the information given in our MT inferences of somebody being not punished (i.e., not-q) is only relevant if there are reasons to assume that the person could have been punished—for example, because the person actually committed the offence. So, if people think that someone actually committed an offence, but find out that the person is not punished, then they should implicitly consider that some exception occurred.

The idea that the negation in MT inferences suggests that the antecedent is actually true, but that some exception happened was already proposed by Oaksford and Chater (2013) and tested empirically by Bonnefon and Villejoubert (2007). To test whether this explanation also applies for legal conditionals, we conducted an online study where participants ($N = 112$) were confronted with negated consequents of legal conditionals (e.g., Sven is not punished for bodily injury) and were asked to write down why somebody would utter this sentence. While in only 7% of the cases participants said that not-q was uttered because of not-p, in 37% of the cases they said not-q was uttered because p happened but some exception occurred. In the remaining cases, participants did not provide any concrete reasons for this utterance, but, for example, described situations when such sentence could be uttered leaving unclear whether they thought p was the case or not (e.g., "after a trial").

Notwithstanding these results, it is still premature to conclude that conversational implicatures are responsible for the high rejection ratings of MT inferences that we found. Further studies are necessary, especially to understand why we did not find any effect of the frequency of exceptions on MT inferences.

A last point we want to make is that participants generated slightly more exceptions in the generation task for conditionals with a low frequency of exceptions than for conditionals with a high frequency of exceptions. Although we did not expect that, this mismatch between amount and frequency of exceptions is not new. For instance, Geiger and Oberauer (2007) show that, although often correlated, the amount of exceptions and the frequency of exceptions do not always lead to the same effects (cf. Fernbach & Erb, 2013). There may be exceptions that happen often (e.g., self-defence or necessity) and exceptions that happen less often

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

(e.g., being coerced by threats to life and physical integrity). It is important to keep this difference in mind especially when dealing with legal conditionals. When asking whether an offender will be punished or not, it is necessary to know how often exceptions occur and not how many different exceptions may exist. We discuss this point further in the General Discussion.

EXPERIMENT 2

In Experiment 1 we showed that people follow their own sense of justice when reasoning with deontic legal conditionals, but that this can be attenuated by changing the modal auxiliary into *will*. In Experiment 2 we tested one further implication of using different modals in legal reasoning. If "should" prompts following one's own deontic preferences, and "will" prompts using one's knowledge about what happens in the real world, then it should be possible to introduce experimental manipulations that affect only inferences for either deontic or factual legal conditionals. Along these lines, in Experiment 2 we presented the offenders as people we love (i.e., family members or best friends). People are usually more forgiving and caring with those they are close with and whom they like (e.g., R. Fehr et al., 2010; McCullough et al., 1998; Sprecher & Fehr, 2005) and do not want them to be hurt or in danger (e.g., Buckley et al., 2012; see also Swann et al., 2014). Consequently presenting the offender as someone close should affect deontic legal conditionals by making people less strict than when deciding about a random person (as they did in Experiment 1). However, it should not affect factual conditionals, because the relationship one has with an offender does not influence what actually happens in the real world (e.g., events involving police, judges, penal code, etc.).

Method

Participants

A total of 42 students participated in the experiment. All participants with academic knowledge about law or formal logic were excluded. The final sample consisted thus of 40 participants (20 female). Half of the participants received deontic legal conditionals, the other half factual legal conditionals.

Material, design, and procedure

Experiment 2 was constructed as Experiment 1, with the only difference that the offender in the second premise was specified as being one's mother, father, best friend (either female—in German: "beste Freundin", or male—in German: "bester Freund"). Which offence was paired with which relative was selected randomly for MP and MT inferences. We made sure that family members and friends were distributed uniformly among all offence categories. As an illustration:

If a person kills another human, then this person should/will be punished for manslaughter.

Your father killed another human.

Should/Will your father be punished for manslaughter?

As in Experiment 1, participants had to solve in addition to the inference task a generation task and a moral outrage task. The offenders in these supplementary tasks were also family members and friends.

Results

Inference task

As in Experiment 1, we transformed the conclusion ratings into rejection ratings and analysed them in a 2 (modal: should vs. will) × 2 (rule acceptance: high vs. low) × 2 (frequency of exceptions: high vs. low) × 2 (inference: MP vs. MT) ANOVA. We found main effects of rule acceptance, $F(1, 38) = 20.93$, $p < .001$, $\eta_p^2 = .355$, and of frequency of exceptions, $F(1, 38) = 32.93$, $p < .001$, $\eta_p^2 = .464$, but also interactions between inference and rule acceptance, $F(1, 38) = 4.74$, $p = .036$, $\eta_p^2 = .111$, between inference, rule acceptance, and modal, $F(1, 38) = 11.51$, $p = .002$, $\eta_p^2 = .232$, and between inference, rule acceptance, and frequency of exceptions, $F(1, 38) = 8.00$, $p = .007$, $\eta_p^2 = .174$. Because of these interactions we continued analysing the data in two separate 2

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

(modal: should vs. will) × 2 (rule acceptance: high vs. low) × 2 (frequency of exceptions: high vs. low) ANOVAs for MP and MT inferences.

For MP inferences we found a main effect of rule acceptance, $F(1, 38) = 22.95$, $p < .001$, $\eta_p^2 = .373$, a main effect of frequency of exceptions, $F(1, 38) = 9.99$, $p = .003$, $\eta_p^2 = .208$, and an interaction between rule acceptance and modal auxiliary, $F(1, 38) = 7.25$, $p = .01$, $\eta_p^2 = .160$. The interaction between modal auxiliary and rule acceptance shows that rule acceptance affected deontic legal conditionals, but not factual legal conditionals: Whereas rejection ratings for deontic conditionals were higher for low ($M = 2.44$, $SD = 1.39$) than for high acceptance rules ($M = 1.04$, $SD = 0.95$), $t(19) = 4.25$, $p < .001$, $d = 1.16$, for factual conditionals this effect did not reach the Bonferroni adjusted alpha level of .025, $t(19) = 2.13$, $p = .046$, $d = 0.44$ ($M = 1.96$, $SD = 0.95$, vs. $M = 1.58$, $SD = 0.79$, respectively). In fact, the rejection ratings we found for deontic legal conditionals were generally higher than the corresponding ones from Experiment 1, showing that participants are tended to punish less when offenders are close relatives. We confirmed this observation by comparing the mean rejection rating of deontic MP inferences in Experiment 2 ($M = 1.74$, $SD = 0.93$) with that in Experiment 1 ($M = 1.06$, $SD = 0.87$), $t(38) = 2.39$, $p = .022$, $d = 0.76$. All other effects, including the interaction between modal auxiliary and frequency of exceptions, were not significant (all Fs ≤ 1.51, $ps \geq .227$).

The ANOVA for MT inferences showed a main effect of rule acceptance, $F(1, 38) = 7.827$, $p = .008$, $\eta_p^2 = .171$, a main effect of frequency of exceptions, $F(1, 38) = 24.22$, $p < .001$, $\eta_p^2 = .389$, and an interaction between both factors, $F(1, 38) = 10.632$, $p = .002$, $\eta_p^2 = .219$. All other effects were not significant (Fs ≤ 1.85, $p \geq .181$). The interaction shows that frequency of exceptions affected conditionals with highly and little accepted rules differently. For highly accepted rules frequency of exceptions did not affect inferences: Participants show low rejection ratings for offences with high ($M = 1.41$, $SD = 1.16$) and low ($M = 1.05$, $SD = 1.47$) frequency of exceptions, concluding that if an offender is not punished (not q), then this offender probably did not commit an offence (not p), $t(39) = 1.49$, $p = .144$, $d = 0.27$. However, for little accepted rules the frequency of exceptions mattered: If the offence had only a low frequency of exceptions, participants showed lower rejection ratings ($M = 1.00$, $SD = 1.15$) than when the offence had a high frequency of exceptions ($M = 2.40$, $SD = 1.82$), $t(39) = 5.81$, $p < .001$, $d = 0.87$ (Bonferroni adjusted alpha level, $\alpha = .025$). For better comparisons we plotted the rejection ratings for deontic and factual legal conditionals together (since we found no effect of modals) and compared them with the corresponding rejection rates for MT in Experiment 1. A 2 (frequency of exceptions: high vs. low) × 2 (rule acceptance: high vs. low) × 2 (experiment: 1 vs. 2) ANOVA confirmed our initial analysis: A main effect of experiment shows that participants indeed accept MT inferences more when the offender is a close relative than when the offender is not specified, $F(1, 78) = 12.66$, $p = .001$, $\eta_p^2 = .14$, and a three-way-interaction between all factors confirms that the interaction between frequency of exceptions and rule acceptance is unique for Experiment 2, $F(1, 78) = 6.316$, $p = .014$, $\eta_p^2 = .075$ (all other effects were explained by the three-way interaction or were not significant).

Generation task

As in Experiment 1, two independent raters coded the amount and quality of exceptions generated ($\rho = .99$ for the amount of exceptions in general, $\rho = .96$ for amount of factual exceptions, and $\rho = .92$ for the amount of deontic exceptions). The amount of exceptions was analysed within a 2 (modal: should vs. will) × 2 (rule acceptance: high vs. low) × 2 (frequency of exceptions: high vs. low) ANOVA. We found only an interaction between modal auxiliary and rule acceptance, $F(1, 38) = 9.48$, $p = .004$, $\eta_p^2 = .2$. Participants in the deontic condition generated slightly more exceptions for low than for high acceptance rules, while participants in the factual condition showed a trend in the opposite direction. However, pairwise t-tests did not reach the Bonferroni adjusted alpha level of $\alpha = .025$ [$t(19) = 2.40$, $p = .027$, $d = 0.32$; $t(19) = 2.09$, $p = .050$, $d = 0.38$, respectively]. Yet, as in Experiment 2, participants in the factual condition generated in 65% of the cases

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

factual exceptions and only in 33% of the cases deontic exceptions. For participants in the deontic condition it was the other way around (75% deontic exceptions and 22% factual exceptions).

Moral outrage

Moral outrage ratings given to offences in the inference task correlated with the corresponding rule acceptance ratings for the same offences from the preliminary study ($r = .91$, $p = .002$ for the correlation between the moral outrage ratings given in the "should" condition, and $r = .93$, $p = .001$ for the ones in the "will" condition).

Discussion

Our results once more show that the modal auxiliary affects the conclusions drawn from legal conditionals. For MP inferences, when asked whether an offender should be punished or not, people generally decide that severe offences should be punished and minor offences not. However, the effect is attenuated when the modal auxiliary is changed to "will". In addition, in Experiment 2 we show that the impact of rule acceptance on punishment decisions also depends on the relationship with the offender. People are less severe when the offender is a close relative, probably because they do not want somebody they like to be punished.

When we started this research we thought that describing offenders to be close people should only affect inferences for deontic legal conditionals, but not for factual legal conditionals. Yet, the present results show that the relationship with the offender also influenced inferences for factual legal conditionals. As shown in Figure 2, different to the results found in Experiment 1, the frequency of exceptions only affected factual legal conditionals with highly accepted rules. When the conditional contained a little accepted rule, participants did not differentiate between offences with high or low frequency of exceptions. On the contrary, they also decided that for offences with a low frequency of exceptions the offender will not be pun-

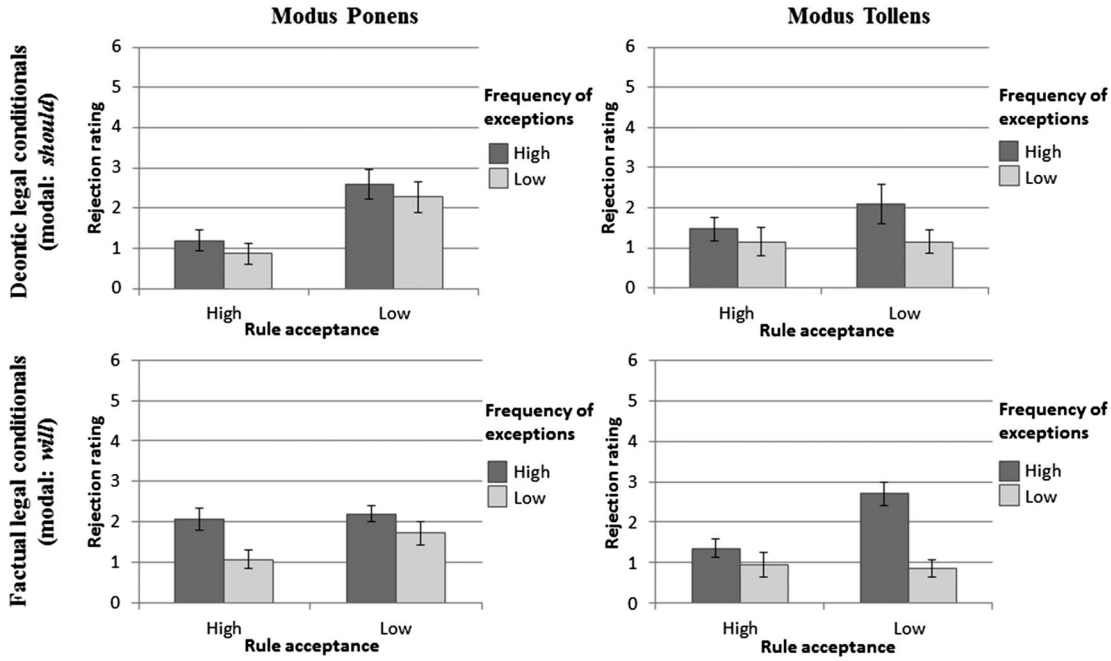

Figure 2. *Rejection ratings (0–6) for modus ponens and modus tollens inferences for deontic and factual legal conditionals in Experiment 2. Error bars represent standard errors.*

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

ished. Although unexpected, this result can be explained if we consider that frequency of exceptions was calculated in the preliminary study of Experiment 1 by asking participants about the frequency of exceptions for an unspecified person ("A person kills another human, but the person is not punished for manslaughter. In how many of 100 cases do you think this is the case?"). It is possible that when participants are told that it was a close relative who committed the offence—and it is a mild and therefore understandable offence—they calculate the probability of being punished differently, assuming that the relative's offence will be one of the few exceptional cases where no punishment follows. As Fernbach and Erb (2013) postulated, the power of an antecedent in predicting the consequent depends on the context. This could also explain why in the generation task we did not find the mismatch between frequency of exceptions and amount of exceptions found in Experiment 1. Given that the frequency of exceptions estimations were calculated in the preliminary study by asking about an unspecified person, and the generation task asked to generate exceptions for one specific close person (e.g., Your best friend killed another person but should/will not be punished for that. Why?), it could be that this mismatch covered any advantage of frequency over amount of exceptions.

Contrary to Experiment 1, participants only seldom rejected MT inferences, suggesting that when the offender is someone close, the negated consequent stated in MT inferences does not suggest the existence of exceptions. We explain these results by remarking that people normally do not believe that beloved people commit regrettable actions. Because of this, when people hear that a beloved person is not punished, they do not conclude that he or she committed an offence and that some exception happened. Instead, they simply conclude that he or she did not commit the offence. Only if the offence is mild, and people know that a lot of exceptions exist, people can imagine that the offence was committed but that an exception occurred (as suggested by the elevated rejection rates for conditionals with little accepted rules and high frequency of exceptions).

GENERAL DISCUSSION

The aim of this research was to investigate whether the modal auxiliary used in the legal conditional influences the conclusions that participants draw. Our results show that for MP inferences it does. On the one hand, if the modal auxiliary is "should", and the inference asks whether an offender should be punished, then the own sense of justice of the reasoner—operationalized in the acceptance of the legal rule—predicts the conclusion that participants choose. On the other hand, if the modal auxiliary is "will", then the frequency of exceptions predicts the conclusions that participants choose, at least when there is no close relationship to the offender. For MT inferences our results were not homogeneous. In Experiment 1, MT inferences were only seldom endorsed, probably because the negated consequent in MT suggested the existence of exceptions. However, in Experiment 2—when the offender was described as a relative or friend—participants endorsed more often the MT inferences concluding that if the offender is not punished, then the offender did also not commit the offence. Only in the case where the offence is mild and susceptible to exceptions do participants conclude that the offender might have committed the offence.

Our results are important for several reasons. First, they show the significance of linguistic factors in reasoning. We showed that the way a conditional rule is phrased affects which kind of information is used for reasoning and which conclusions are drawn. We are not the first ones to highlight the importance of linguistic factors in the psychology of reasoning. For example, Schmeltzer and Hilton (2014) showed that the pragmatic implications of the antecedent influence the conclusions that participants draw. Similarly—and also in accordance with our results—Bonnefon and Villejoubert (2007) showed that the negated consequent in MT inferences suggests the existence of exceptions. There are many different kinds of conditionals in the literature—for example, causal conditionals, conditionals describing threats, tips and promises (see Dieussaert, Schaeken, & d'Ydewalle, 2002), or consequential conditionals

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

describing outcomes (Bonnefon & Hilton, 2004), making clear how important it is to track their different formulations when making predictions about how people reason with them. The distinction between deontic and factual conditionals is especially relevant. As we have shown for legal reasoning, deontic and factual conditionals activate different system of rules: While the former activates the moral system, the latter activates the knowledge about the real world. Some might argue that our factual conditionals were still deontic in certain way because they still represented some kind of rule. We agree with that. However, instead of weakening our arguments, this criticism supports our hypothesis about the relevance of modals in reasoning: Although the content was in both cases somewhat deontic, the different modals used in order to express the rule made participants draw different conclusions.

Second, the results of Experiment 1 show that the best predictor for factual legal conditionals was the frequency of exceptions and not the amount of different exceptions per se. Although in some studies merely the amount correctly predicted the inferences that participants draw (e.g., De Neys et al., 2003a), this seems not to be always the case. In legal reasoning, when trying to predict whether a person will be punished or not, it matters how probable this is—in other words, how often this will happen—and not how many different causes may exist in principle for not punishing somebody. As Fernbach and Erb (2013) proposed, the power of the antecedent in predicting the consequent does not only depend on the amount of exceptions, but also on how relevant they are, and in this way—we think—also on how often they occur. The amount of generated exceptions is not always a good predictor because —when measured as in most studies by giving participants 1.5 minutes to generate—we do not know how often participants think each exception occurs or how easy it was for them to retrieve them. In generation tasks we can say that the first generated exception is probably also the most frequent one for this very specific conditional; however, we cannot compare the frequency of exceptions across conditionals. It is necessary to conduct further studies to disentangle the relative importance of these factors in order to be able to understand their relationship and influence on conclusions. If further studies corroborate the results found in this paper, then in a next step we could let participants assign probabilities to the premises and conclusions of legal conditionals and test how far probabilistic accounts can explain our results (cf. Evans, Thompson, & Over, 2015; Oaksford & Chater, 2007; Singmann et al., 2014).

Third, our results support recent studies on the "new paradigm" of cognitive psychology showing that the distinction between reasoning and decision making is artificial (Evans, 2012). The fact that the acceptance of the rule—computed by the preference for different outcomes—and the relationship to the offender influenced participants' conclusions shows how strongly interrelated decision making and reasoning are. Participants use their subjective gains and losses in reasoning tasks to decide whether they agree or not with a given conclusion. It may be that in former studies, focused on "disinterested" conditionals (Bonnefon, Girotto, & Legrenzi, 2012; p. 28), utilities might not have played such a big role, but nowadays in the new psychology of reasoning—which deals with everyday reasoning—utilities are important (e.g., Bonnefon, 2009, 2012; Bonnefon et al., 2012; Oaksford & Chater, 2007). This is the case especially when dealing with conditionals of high societal relevance such as legal conditionals. For instance, one could try to implement Oaksford's and Chater's (2007) decision theoretic approach on deontic reasoning in our tasks. Although Oaksford and Chater (2007) apply their approach mainly to the deontic selection task, the idea of assigning utilities to the different outcomes of a rule can be interesting also for our studies. This is especially considering the distinction that Oaksford and Chater make with respect to factual and deontic conditionals: Whereas reasoning with the former can be explained by probabilities, the latter can be explained by utilities (Oaksford & Chater, 2007, 2009; see also Over et al., 2004).

One point that needs further investigation is the results for the MT inferences. In Experiment 1 we argued that according to the principle of relevance

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

(Sperber & Wilson, 1995) the categorical of MT inferences (in our case: "The person should/will not be punished for offence X") makes people aware of exceptions. The utterance denying that person is punished is only relevant if there are reasons for thinking that the offender could be punished but some exception occurred (e.g., not being caught, having some legal justifications). If this conversational implicature (see Grice, 1975) of legal MT inferences is correct, then it is necessary to (a) reinterpret existing data on MT endorsement, and (b) conduct further studies to identify moderators of this effect. According to Experiment 2, a possible moderator could be a person's personal commitment to the conditional. The denial of the consequent enhances people's thought of exceptions only if there are no personal pragmatic implications in conflict with it. When people do not want to believe that the antecedent is true (here: someone close doing something they dislike) then hearing that the consequent is not the case fulfils their expectations, making them conclude that the antecedent is not true. In other words, the motivation to conclude that the antecedent is not the case might bring people to ignore exceptions and thereby endorse the classically correct conclusion to MT. All in all, further studies are necessary to fully understand the role of conversational implicatures on reasoning and especially on MT.

Another task for the future is to investigate the influence of other modals on legal reasoning. For example, the modal "may" could make people more liberal given that rules containing such *permission* modals are often labelled as weak rules (Beller, 2008). On the contrary, strong rules as such containing the necessary modal "must" (Beller, 2008) could make people more conservative. It would also be interesting to test whether the effect of phrasing also applies in other domains besides legal reasoning. We can imagine modals also influencing reasoning with everyday rules like the rules a mother gives to her child (e.g., "If you tidy your room, you may go outside to play"; see Manktelow & Fairley, 2000; Manktelow & Over, 1991) or rules from a work environment (e.g., "If an employee works on the weekend, then that person gets a day off during the week"; see Gigerenzer & Hug, 1992; Kilpatrick, 2009).

Finally, in further studies the consequent of the legal conditionals could be made more precise. In the present study we always used the word "punishment" in the consequent, which can have different meanings, from a fine, to life imprisonment, or even to the death penalty in some countries. We can imagine that pairing legal conditionals with very lenient or very harsh consequents may affect inferences. For instance, if a deontic legal conditional is paired with a too lenient consequent (e.g., "If a person kills another human, then this person should be punished with a fine"), or with a too harsh one (e.g., "If a person steals, then this person should be punished with life imprisonment") people might refuse to draw MP conclusions although they are against the committed offence. The same could happen with factual legal conditionals: If the consequent does not describe what is believed to be the usual punishment for a given offence, then people might also refuse to endorse MP.

In sum, our results corroborate the complexity of human reasoning. Contrary to the assumptions of classical logic, the way people interpret and reason with conditionals often depends on its content, on the personal relevance the rule has, and on its linguistic phrasing. In some cases people may consider their own inner values when drawing conclusions; however, in other cases factual information about exceptional situations can be more important and take precedence. It is necessary to keep in mind these differences when trying to predict how people understand and hence what they infer from if–then conditional rules.

REFERENCES

Alter, A. L., Kernochan, J., & Darley, J. M. (2007). Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law and Human Behavior*, *31*, 319–335.

Bäcker, C. (2009). Der Syllogismus als Grundstruktur des juristischen Begründens? *Rechtstheorie*, *40*, 404–424.

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

Bäcker, C. (2010). Rules, principles, and defeasibility. In M. Borowski (Ed.), *On the nature of legal principles, ARSP-Beiheft 119* (pp. 79–91). Stuttgart: Franz Steiner Verlag.

Baratgin, J., Over, D. E., & Politzer, G. (2014). New psychological paradigm for conditionals and general de finetti tables. *Mind & Language*, *29*, 73–84.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science*, *22*, 25–51.

Beller, S. (2008). Deontic norms, deontic reasoning, and deontic conditionals. *Thinking and Reasoning*, *14*, 305–341.

Beller, S. (2010). Deontic reasoning reviewed: Psychological questions, empirical findings, and current theories. *Cognitive Processing*, *11*, 123–132.

Bonnefon, J. F., (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review*, *116*, 888–907.

Bonnefon, J-F., (2012). Utility conditionals as consequential arguments: A random sampling experiment. *Thinking & Reasoning*, *18*, 379–393.

Bonnefon, J-F., Girotto, V., & Legrenzi, P. (2012). The psychology of reasoning about preferences and unconsequential decisions. *Synthese*, *185*, 27–41.

Bonnefon, J-F., & Hilton, D. J. (2004). Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 28–37.

Bonnefon, J-F., & Vautier, S. (2010). Modern psychometrics for the experimental psychology of reasoning. *Acta Psychologica Sinica*, *42*, 99–110.

Bonnefon, J-F. & Villejoubert, G. (2007). Modus Tollens, Modus Shmollens: Contrapositive reasoning and the pragmatics of negation. *Thinking & Reasoning*, *13*, 207–222.

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

Bucciarelli, M., & Johnson-Laird, P. N. (2005). Naïve deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*, *50*, 159–193.

Buckley, L., Chapman, R. L., Sheehan, M. C., & Cunningham, L. (2012). Keeping friends safe: A prospective study examining early adolescent's confidence and support networks. *Educational Studies*, *38*, 373–381.

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

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. *Advances in Experimental Social Psychology*, *40*, 193–236.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*, 284–299.

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

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391–416.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

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.

Darley, J. M. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science*, *5*, 1–23.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*, 324–336.

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

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.

Edgington, D. (1995). On conditionals. *Mind*, *104*, 235–329.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning*, *19*, 249–265.

Elqayam, S., Thompson, V. A., Wilkinson, M. R., Evans, J. S., & Over, D. E. (in press). Deontic

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016

introduction: A theory of inference from is to ought. *Journal of Experimental Psychology: Learning, Memory and Cognition*. doi:10.1037/a0038686

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*, 978–996.

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, England: Oxford University Press.

Evans, J. S. B. T., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00398

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin*, *136*, 894–914.

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, *58*, 609–626.

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1327–1343.

Gazzo Castañeda, L. E., & Knauff, M. (2015). *Defeasible reasoning with legal conditionals*. Manuscript submitted for publication.

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

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*, 127–171.

Groefsema, M (1995). Can, may, must and should: A relevance theoretic account. *Journal of Linguistics*, *31*, 53–79.

Green, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.

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: Academic Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998–1002.

Johnson-Laird, P. N. (1978). The meaning of modality. *Cognitive Science*, *2*, 17–26.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hove, UK: Erlbaum.

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

Keller, L. B., Oswald, M. E., Stucki, I., & Gollwitzer, M. (2010). A closer look at an eye for an eye: Laypersons' punishment decisions are primarily driven by retributive motives. *Social Justice Research*, *23*, 99–116.

Kilpatrick, S. G. (2009). *Pragmatic factors of deontic reasoning* (Doctoral dissertation). University of Wolverhampton). Retrieved from http://wlv.openrepository.com/wlv/handle/2436/71021

Kilpatrick, S. G., Manktelow, K. I., & Over, D. E. (2007). Power of source as a factor in deontic inference. *Thinking & Reasoning*, *13*, 295–317.

Leiner, D. J. (2014). *SoSci Survey (Version 2.4.00-i)* [Computer Software]. Retrieved from https://www.soscisurvey.de

Manktelow, K. I., & Fairley, N. (2000). Superordinate principles in reasoning with causal and deontic conditionals. *Thinking & Reasoning*, *6*, 41–65.

Manktelow, K. I., & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition*, *39*, 85–105.

Markovits, H., & Potvin, F. (2001). Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals. *Memory & Cognition*, *29*, 736–744.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L. Jr, Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, *75*, 1586–1603.

Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning*, *1*, 121–152.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*, 349–357.

Oaksford, M., & Chater, N. (2003). Probabilities and pragmatics in conditional inference: Suppression and order effects. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on*

reasoning, judgment and decision making (pp. 95–122). Chichester, UK: John Wiley & Sons Ltd.

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

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, *32*, 69–84.

Oaksford, M., & Chater, N. (2010). Cognition and conditionals: An introduction. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 3–36). Oxford: Oxford University Press.

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

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 883–899.

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

Over, D. E., Manktelow, K. I., & Hadjichristidis, C. (2004). Conditions for the acceptance of deontic conditionals. *Canadian Journal of Experimental Psychology/ Revue canadienne de psychologie expérimentale*, *58*, 96–105.

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Reasoning with exceptions: An event-related brain potentials study. *Journal of Cognitive Neuroscience*, *23*, 471–480.

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

Quelhas, A. C., Byrne, R. M. J. (2003). Reasoning with deontic and counterfactual conditionals. *Thinking and Reasoning*, *9*, 43–65.

Schmeltzer, C., & Hilton, D. J. (2014). To do or not to do? A cognitive consistency model for drawing conclusions from conditional instructions and advice. *Thinking & Reasoning*, *20*, 16–50.

Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, *5*, 316. doi:10.3389/fpsyg.2014.00316

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

Sprecher, S., & Fehr, B. (2005). Compassionate love for close others and humanity. *Journal of Social and Personal Relationships*, *22*, 629–651.

Swann, W. B., Gómez, A., Buhrmester, M. D., López-Rodríguez, L., Jiménez, J., & Vázquez, A. (2014). Contemplating the ultimate sacrifice: Identity fusion channels pro-group affect, cognition, and moral decision making. *Journal of Personality and Social Psychology*, *106*, 713–727.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281.

Weidenfeld, A., Oberauer, K., & Hörnig, Robin (2005). Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *58*, 1479–1513.

Downloaded by [Universitaetsbibliothek Giessen] at 01:25 12 April 2016