

Reasoning with Fundamental Rights

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

Antje Stemmler (Antje.Stemmler@psychol.uni-giessen.de)

Markus Knauff (Markus.Knauff@psychol.uni-giessen.de)

Justus-Liebig University, Experimental Psychology and Cognitive Science, Otto-Behagel-Str. 10F,
35394 Giessen, Germany

Abstract

People often withdraw previously drawn conclusions in light of new information. This defeasible reasoning is also important for law, where judges often have to change their verdicts in light of new evidence. Here we investigate defeasibility in the context of conflicting fundamental rights. When, for instance, law to property conflicts with law to information, can one of these rights be “defeated” by the other? We embedded conflicting fundamental rights in inference tasks (Experiment 1) and in elaborated vignettes (Experiment 2). Results show that people decide between two conflicting fundamental rights in a rational way. Case by case, participants protected that fundamental right whose violation evoked the highest moral outrage (Experiment 1) or whose violation was considered to be more serious (Experiment 2). We discuss the implications of our findings for law theory and psychology.

Keywords: defeasibility, legal reasoning, conditionals

Introduction

Are humans rational? This question has concerned psychologists and philosophers for a long time. Philosophers have developed norms for rational thinking and psychologists have tested them empirically. In many of these experiments, classical logic was used as a norm for rationality. Participants were confronted with inference tasks, consisting of a conditional and a fact, and asked to indicate what follows necessarily. One example is Modus Ponens (MP):

If Ann is hungry (p) then she gets something to eat (q).

Ann is hungry (p).

Ann gets something to eat (q).

MP is a valid inference because in classical logic the antecedent (p) is sufficient (but not necessary) for the consequent (q) (e.g., Thompson, 1994; 1995). Yet, many participants made logical errors in such inference tasks, rejecting otherwise valid conclusions. Nowadays, however, it is known that these “errors” are only a consequence of the complexity of human everyday reasoning. In everyday situations, many factors that are irrelevant for classical logic have to be considered and weighted in order to arrive at a reasonable conclusion. For instance, if Ann is on a strict diet, it may be rational to conclude that she will not get something to eat even if she is hungry. Contrary to classical logic, where no additional information can make a conclusion false, everyday reasoning is non-monotonic and defeasible (e.g. Oaksford & Chater, 1995; 2013; Stenning & van Lambalgen, 2005).

This phenomenon called defeasibility is also very important in legal reasoning (e.g., Bäcker, 2010; Prakken &

Sartor, 2004). Judges are often confronted with complex cases, in which they have to arrive at rational verdicts. At first sight, we might thus think that classical deduction is an appropriate norm for legal reasoning. For instance, considering that the penal code says that “If a person kills another human, then the person has to be punished for manslaughter” we can conclude from the fact that a person killed another human that the person has to be punished for manslaughter. However, in many cases there are exculpatory circumstances that make it rational to reject this conclusion. There are different exculpatory circumstances defined in penal code, such as self-defense, necessity or psychological disorders. In light of those circumstances, judges know that the otherwise valid conclusion of punishment has to be defeated in favor of acquittal. But what happens if there are no clear rules on how to reason? Although this might sound counterintuitive for legal contexts, this nonetheless happens in federal constitutional courts, when two fundamental rights are in conflict. Imagine, for instance, that you live in a foreign country and the only way to hear news from your hometown is to mount a parabolic antenna on the facade of your rented flat. The landlady nonetheless prohibits it to you. Your right to receive information thus conflicts with the property law of the landlady. In general terms, all fundamental rights are equally important and have to be granted. So, which fundamental right has to be preferred over the other? Can one fundamental right “defeat” the other?

The aim of this paper is to investigate how people reason with conflicting fundamental rights. For this, we will first discuss the psychological literature on defeasible reasoning and then the law theoretic framework on fundamental rights.

Withdrawing from valid conclusions

Many factors influence defeasible reasoning. One important factor is background knowledge. Just as shown in the introduction of this paper, when people know circumstances that prevent the consequent to occur although the antecedent is true, they reject otherwise valid conclusions. These circumstances are often called disablers or defeaters. But not only the availability of defeaters is important, also is their amount (e.g., Cummins, 1995; De Neys, Schaeken, & d’Ydewalle, 2003a), their relative strength (De Neys, Schaeken, & d’Ydewalle, 2003b), and their frequency of occurrence (Geiger & Oberauer, 2007). The more defeaters there are, the more associated or salient these defeaters are, and the more often they occur, the more readily a conclusion is withdrawn. However, another important factor are utilities. Bonnefon (2009; Bonnefon & Hilton, 2004) showed

that when people make inferences about actions, they consider the costs and benefits of this action given the consequences. For instance, when presented with the conditional “If Mary’s TV is broken, she will have it fixed” and the fact that “Mary’s TV is broken” participants refuse to conclude that Mary will have their TV fixed when presented with the additional information that “If Mary has her TV fixed, she will not be able to pay the electricity bill”. The consideration of utilities during reasoning illustrates the closed connections between reasoning and decision making. When people reason in their daily lives, it is not just for the sake of reasoning per se, but to reach a goal – may this goal be something simple like getting a TV fixed or something relevant for society like reaching a legal verdict. Because goals are also relevant in law we expect utilities to play a similarly important role in legal reasoning.

Legal reasoning

Law is defeasible in several ways (cf. Prakken & Sartor, 2004): during police investigations when new evidence “defeats” previous insights, during trials when attorneys and prosecutors defeat each other’s arguments, and in the application of legal rules in light of exculpatory evidence. The role of utilities in the application of legal rules has been tested by Gazzo Castañeda and Knauff (2016). In several experiments, laypeople and lawyers were confronted with legal conditionals embedded in MP inferences, which were presented together with exculpatory evidence (e.g., “If a person kills another human, then the person should be punished for manslaughter; Bob killed another person; Bob is schizophrenic and had a delusion of an attack against him; Should Bob be punished for manslaughter?”). As expected, lawyers considered exculpatory circumstances as prescribed by the penal code and irrespective of how morally outraging the offence was, deciding not to punish in light of exculpatory circumstances. Laypeople, instead, had difficulties in accepting exculpatory circumstances when the offence was highly morally outraging (e.g., maltreatment of wards), but not if the moral outrage was only low (e.g., illegal gambling). We argued that utilities might be responsible for this moral outrage effect. People can only feel secure in a society where they can be sure that the important rules are respected and offenders punished. The benefit of saving one’s own feeling of security is thus weighted more than the costs of punishing somebody erroneously. This overweighing of one’s own feeling of security is known from the belief in a just world literature (see Lerner, 1970), where people even tend to blame the victims of offences only to preserve their belief that people get what they deserve and that bad things only happen to bad people. From a utilitarian point of view, the punishment of offenders is thus of high utility – and the higher the moral outrage, the higher this utility is. Is it therefore possible that moral outrage also affects the weighing of fundamental rights?

Fundamental rights are generally coded in the constitution. The most known examples are right to dignity, liberty, freedom of thought and of expression, or right of property.

All of these have to be respected and protected. However, there are instances when two or more fundamental rights are in conflict, such as in the introductory example when the right to information conflicts with right to property. Judges in the federal court are thus faced with the problem that they have to decide which one deserves more importance, although both are theoretically equally important. This weighting of fundamental rights is called *balancing* and is an important case by case decision with no clear rules on how to decide. That is, cases with the same conflicting fundamental rights can (and should) end up with different verdicts due to case-specific details. Because of these missing rules, some law theorists argue that balancing cannot be rational (e.g., Habermas, 1992). Alexy (2003), however, argues that balancing can be rational by comparing for every single case the detriment of one fundamental right i with the importance of satisfying the other fundamental right j . This is done by the so-called weight formula, which computes the ratio between the case-specific weights I_i and I_j ($G_{ij} = I_i / I_j$). The first weight I_i stands for the violation intensity of fundamental right i by protecting fundamental right j , and the second weight I_j stands for the importance of protecting fundamental right j by violating fundamental right i . Applied to our concrete example, this would result in the following two questions: How serious is the invasion of the right to information by prohibiting the installation of the parabolic antenna? How important is it to protect the right to property by prohibiting the installation of the parabolic antenna? Already Darley showed in several experiments that the perceived seriousness or severity of offences is highly correlated by moral outrage (e.g., Alter, Kernochan, & Darley, 2007; Carlsmith, Darley, & Robinson, 2002; Darley, Carlsmith, & Robinson, 2000). Therefore, we also expect that the case-specific weights of fundamental rights will depend on moral outrage: If the invasion of fundamental right A is considered more morally outraging than the invasion of fundamental right B, then fundamental right A should be protected over B.

In this paper, we combine the domains of defeasible reasoning from psychology with the concept of balancing from legal theory. In Experiment 1, we embedded two fundamental rights into conditional reasoning tasks and asked participants what should follow. In Experiment 2, we embedded conflicting fundamental rights into longer vignettes and asked participants for the case-specific weights I_i and I_j .

Experiment 1

Methods

Participants We tested 40 people (21 male) without legal expertise. They were on average 26.62 years old ($SD=6.93$).

Material We took 16 real conflicts of fundamental rights from the German constitutional court and embedded them in defeasible inference tasks. Each problem started with a conditional containing one fundamental right A. Next, we presented a concrete situation as second premise in which the fundamental right A was involved, followed by a third

premise in which the fundamental right A is applied to this concrete situation (MP). Then, the second fundamental right B was presented as a defeater that is in conflict with the previous information. Finally, the conclusion was presented as a question asking either for the application of fundamental right A (Example 1) or fundamental right B (Example 2):

If a person’s personal security is endangered, then its protection has to be warranted.

Person A’s house has to be searched and seized because A is suspected to have death threatened person B.

To protect B this search and seizure can be authorized.

The suspect A has nonetheless right to privacy.

Should the house of suspect A be searched and seized?

If the personality rights are in danger, then their protection has to be warranted.

A celebrity is photographed without permission.

Due to the personality rights, all people’s privacy has to be protected.

The press has nonetheless right to freedom of the press.

Should the celebrity be photographed by the press without permission?

Participants gave ratings from 1 (not at all) to 7 (definitely). Therefore, the higher a rating was, the more the participants preferred one fundamental right over the other. We refer to this as “preference rating”.

We created two versions of the experiment by changing the order of the fundamental rights A and B (version 1 and 2). If in one version one fundamental right was presented as the conditional and the other as the defeater, then in the other version it was the other way around. The conclusion, however, always asked for the same fundamental right. This allowed us to control for order effects.

To measure moral outrage we conducted a norming study in which participants ($N=34$) rated on a seven point Likert-scale how morally outraged they would feel if the fundamental rights from the inference tasks were violated (e.g., “How outraging do you find it when a celebrity is photographed without permission?”). Because in each conflict situation there were two fundamental rights involved, this resulted in 32 violation ratings, ranging from 2.65 to 6.09.¹

Procedure and Design The experiment was programmed on Superlab 4.5. Participants were tested individually on a desktop computer and were instructed that no right or wrong answers exist. The instructions included one practice problem. All 16 problems were presented randomly and separated by fixation crosses. The single premises were presented sequentially on separate screens. Participants could switch to the next premise by pressing the space bar. The last premise was always the question about the conclusion. It was

written in red font and was presented together with the 7-point-Likert scale. The experiment was thus one factorial with “version” as a between subject variable.

Results

Comparisons between the two versions of the experiment revealed no differences in preference ratings, $t(38)=1.36$, $p=.181$. That means that regardless of whether a fundamental right was presented as the conditional or as the defeater, this did not affect its evaluation in the conclusion. This allowed us to test the effect of moral outrage on preference ratings. For this, we first compared the two fundamental rights in each problem on the basis of the moral outrage ratings they got in the norming study. We looked at which fundamental right violation got higher moral outrage ratings and should thus be preferred. These predictions were then compared with the actual preference ratings participants gave in the experiment. Mean preference ratings over 4 (i.e., the scale midpoint) were classified as in favor, and ratings below 4 against the fundamental right presented in the conclusion (no mean preference rating = 4). Descriptively, the moral outrage ratings allowed us to correctly predict 11 out of the 16 conflict situations. To corroborate this statistically, we tested the preference ratings of each inference task against 4 with one sample t-tests and a Bonferroni adjusted alpha of 0.0031. Results are in Table 1. Of the 16 comparisons, 6 were not significantly different from 4, meaning that participants were neither in favor nor against the fundamental right presented in the conclusion. From the remaining 10 problems, however, we were able to predict statistically 8 conflict situations.

Table 1: Predicted and actually preferred fundamental rights. Predictions were based on the moral outrage (MO) ratings from the norming study. Preference ratings of the actually preferred rights were tested against the scale midpoint 4 (Sign., Bonferroni adjusted alpha: 0.0031).

Item	MO of A	MO of B	Predicted	Preferred	Sign.
1	3.15	5.44	B	A	<.001
2	4.88	5.59	B	A	<.001
3	3.88	3.32	A	A	.001
4	3.85	4.21	B	B	.001
5	3.71	4.59	B	A	.147
6	3.29	4.71	B	B	<.001
7	5.79	4.38	A	A	<.001
8	3.76	5.26	B	A	.008
9	4.41	5.06	B	B	.386
10	3.24	5.53	B	B	<.001
11	3.82	3.74	A	A	.103
12	5.76	4.06	A	A	<.001
13	5.18	5.44	B	B	.309
14	2.65	6.09	B	B	<.001
15	3.39	3.26	A	B	.305
16	5.32	5.29	A	A	<.001

As an additional measure of the relevance of moral outrage for balancing, we took the moral outrage ratings from the norming study and used these ratings to classify the fundamental rights in each problem as either low (ratings from 2.65 to 3.76), medium (from 3.82 to 4.88), or high (from 5.06 to 6.09) morally laden (the cut offs resulted from the division of our fundamental rights into these three groups).

¹We have the original raw data file from the norming study, but, alas, have lost the handwritten surveys on paper. We therefore conducted a second norming study with the same materials. The results were similar and so we used the data from the original norming study. Interestingly, the few items where we found slight differences were the ones related to immigration – a topic that recently became highly controversial in Germany and many other countries.

An analysis of the respective preference ratings showed that when high and medium morally laden fundamental rights were in conflict, participants gave higher preference ratings for conclusions asking for the highly morally laden right ($M=5.39$; $SD=1.20$) than for conclusions asking for the medium morally laden right ($M=4.46$; $SD=1.36$), $t(39)=3.06$, $p=.004$. The same was the case for problems where medium and low morally laden fundamental rights were in conflict: conclusions asking for the medium morally laden right got higher preference ratings ($M=4.5$; $SD=1.89$) than their counterpart ($M=3.25$; $SD=1.01$), $t(39)=3.72$, $p=.001$. We also found the same pattern for problems with high and low morally laden fundamental rights, but the difference in preference ratings ($M=4.15$; $SD=1.09$ vs. $M=3.73$; $SD=1.36$) did not reach significance, $t(39)=1.38$, $p=.176$ (Bonferroni adjusted alphas: 0.0167).

Discussion

Moral outrage was an important predictor when deciding between two fundamental rights in a defeasible reasoning paradigm. Participants protected more often that fundamental right whose violation provokes the highest moral outrage. A fundamental right was therefore only considered as a defeater if its violation was morally outraging enough.

Our results demonstrate the defeasibility of human reasoning. Even in contexts where we expect people to reason deductively – such as in law – reasoning is often defeasible. In fact, it is difficult to describe balancing through deduction. Deduction would imply that it should be (in principle) possible to enumerate all defeaters beforehand as part of the antecedent (e.g., If right to property is in danger and it does not conflict with right to information, then we have to protect it). Yet, this is not possible for balancing because cases with the same conflicting fundamental rights can end up with different verdicts due to case-specific details. An important task for cognitive psychologists therefore is to understand the cognitive processes behind balancing. Our research is a first step in this direction.

An open question is, however, whether participants are indeed capable to consider these case-specific details during balancing. Is balancing just a theoretical concept from legal theory? Or do people in fact balance and defeat fundamental rights differently depending on case-specific details? We tested this in Experiment 2.

Experiment 2

In Experiment 2 we used a new experimental paradigm. First, we embedded the conflicting fundamental rights in elaborated vignettes. The vignettes described many case-specific details and were thus more realistic than the inference tasks from Experiment 1. Second, half of our participants were people with legal expertise. This allowed us to investigate balancing in a more realistic “court like” setting.

Methods

Participants We tested 40 laypeople (17 male) and 40 lawyers (already graduated ones and advanced law students; 18

male). On average, laypeople were 24.3 years old ($SD=4.2$; one missing value) and lawyers 24.8 years old ($SD=3.1$).

Material We constructed our material by summarizing 8 real cases from the German federal constitutional court (i.e., BVerfGE), and embedding these into vignettes. Each vignette contained case-specific details such as the matter of facts, the case history (e.g., accusations, levels of jurisdiction involved), and the parties’ arguments in favor or against the different fundamental rights. We selected our cases in such a way that two of them always contained the same conflicting fundamental rights, but received different verdicts from the constitutional court. These final verdicts were, however, not included in the vignettes. We had thus four pairs of cases: two cases of right to information vs. right to property, two cases of personality rights vs. right to freedom of press, two cases of personality rights vs. right to freedom of speech, and two cases of right to bodily integrity vs. the public interest to legal action. The vignettes were 324 to 506 words long and were developed by an advanced law student with the supervision of an experienced legal researcher.

The participants’ tasks were (1) to come to a final verdict and (2) to determine the specific weights for I_i and I_j . The question about the final verdict was formulated according to the legal theoretic tradition: “How would you decide? Which interest should resign: [Right A] or [Right B]?” Participants could select between “[Right A] should resign”, “[Right B] should resign”, and “Both interest deserve equal protection (standoff)”. The question about the specific weights was split into two parts: First, participants had to judge the intensity of violation of right A (e.g., “How intense do you think is the violation of T’s right to information by prohibiting him to install a parabolic antenna?”). Second, participants had to judge the importance of protecting right B (e.g., “How important is it to protect the right to property of the landlady by prohibiting the installation of the parabolic antenna?”). Participants had to select between “little”, “medium”, “very”.

Procedure and Design The experiment was conducted via paper and pencil. Each vignette was presented on a small booklet containing on the first page the vignette and on the second page the questions about (1) the verdict and (2) the specific weights (in this order). Participants were instructed to imagine they were judges in the constitutional court. Each participant received 4 of the 8 vignettes, one of each pair. The order of the vignettes was randomized. The experiment thus followed a 4(type of conflict) x 2(pair) design, with “pair” as a between subjects variable and the type of conflict as a within subjects variable.

Results

We first analyzed in how many cases the specific weights predicted the overall verdicts. As a correct prediction we counted (1) cases in which the fundamental right protected in the final verdict was also the one with the highest specific weight, and (2) cases in which participants weighted both rights equally in the questions about the specific weights and selected “standoff” as the verdict. This analysis showed

that correct predictions were significantly above chance: we could predict 71.9% of the laypeople's, $t(39)=5.31$, $p<.001$, and 81.9% of the lawyers' verdicts, $t(39)=8.64$, $p<.001$.

In a second step, we analyzed whether participants considered the case-specific details. Therefore, we looked at the four pairs of conflicting fundamental rights and compared within each pair how often *Right A*, *Right B*, or standoff were selected. We compared the frequency distributions of the three kinds of verdicts with Freeman-Halton tests. Indeed, results showed that in light of different case-specific details, participants gave different verdicts for the same conflicting fundamental rights. Laypeople did so for 2 of the 4 pairs of cases (personality rights vs. freedom of speech: $p=.002$; bodily integrity vs. public interest: $p=.001$), and lawyers did so for 3 of the 4 pairs of cases (personality right vs. freedom of speech: $p=.009$; personality rights vs. freedom of press: $p=.002$; bodily integrity vs. public interest: $p<.001$). However, only in 38% (laypeople) and in 55% (lawyers) of the cases the participants' final verdicts was the same as the actual verdicts from the constitutional court.

Discussion

Experiment 2 shows that participants often decide between conflicting fundamental rights by considering case-specific details. This supports our main assumption that balancing is defeasible. In our study, participants did not apply some general rule (e.g., right to information deserves more importance than right to property), but decided on a case by case manner whether a specific fundamental right counts as a defeater or not. This defeasibility seems to be well-captured by Alexy's (2003) weight formula. An interesting question now is whether the basic idea of this formula is also helpful for understanding defeasibility outside the legal context. Take our initial example of Ann being hungry. Maybe people decide to defeat the conclusion that people eat when they are hungry by comparing the weights of "if hungry then eating" and "if hungry then not eating". Interestingly, this comparison of weights is similar to the concept of conditional probabilities (e.g., Evans & Over, 2004). Many theories on conditional reasoning argue that defeasibility results from the fact that conditionals are understood as the conditional probability $P(q|p)$, which is computed by dividing $P(p\&q)$ with $P(p\&q)+P(p\&\text{not-}q)$. That is, similar to the weight formula, the weight (here the probability) of p and q is compared with the one of p and $\text{not-}q$. This similarity, we think, deserves more attention from psychology and also from law theory.

Another point that also deserves attention is the mismatch between the final verdicts of our participants (laypeople and lawyers) and the actual verdicts of the constitutional court. On the one hand, participants followed the weight formula. Thus, they weighted the single fundamental rights in a rational way. On the other hand, our results indicate that they used specific weights that differed from those used by the constitutional court. This might be a result of our specific task setting, relatively low test power, the small number of vignettes we used, and the limited ecological validity of our

study. However, another interpretation is that the ethical values and moral principles that drive people's decisions differ from that of our legal system. We think that this is an important research topic at the intersection of cognitive science, social psychology, legal theory, and moral philosophy.

General Discussion

We used methods from cognitive psychology to investigate the concept of *balancing* from legal theory. Our results show that people are willing to defeat single fundamental rights if they are in conflict with other fundamental rights. This defeasibility happens in a case-specific manner and not only when conflicting fundamental rights were presented in inference tasks, but also when they were embedded in ecologically more valid vignettes.

Our findings are important for several reasons. First, they show the importance of defeasible reasoning in many areas of real life. Defeasibility is important if we judge how severe violations of fundamental rights are and when we weight the importance of conflicting fundamental rights. Interestingly, however, some law theorists do not consider balancing as defeasible. According to Bäckér (2010), defeasibility describes the capacity to accommodate legally relevant exceptions. Therefore, only "normal" legal rules would be defeasible (e.g., those from penal code), but not fundamental rights. Fundamental rights are legal principles that have to be optimally achieved taking into account all possible circumstances, including other conflicting rights. Therefore, one fundamental right cannot be an exception to another fundamental right (Bäckér, 2010). From a psychological perspective two reasons speak against this view. The first is an empirical: the important aspect of psychological defeasibility is that people change their conclusions in light of new evidence and this certainly happens when one decides against one right in light of another right. The second reason is a theoretical: that one fundamental right cannot be an exception to another one does not speak against the defeasibility of balancing. In fact, the case by case weighting of fundamental rights is precisely what makes balancing defeasible and non-monotonic. Would one fundamental right be considered an "exception" of another, then we could theoretically enumerate it as part of some rule and reason deductively.

Second, our findings also help to understand the psychological variables behind the weight formula. According to Alexy (2003), balancing depends on specific weights, which reflect how serious it is not to protect one right or the other. It is, however, not clear how exactly judges determine this "seriousness". We operationalized this seriousness through moral outrage, which resulted to be a good predictor for the final verdicts. Is it thus possible that judges' balancing of fundamental rights is influenced by the level of moral outrage? The fact that both, moral outrage (Experiment 1) and the specific weights (Experiment 2), were good predictors for the final verdicts suggests this. Certainly, most judges and lawyers will not accept this view and it is indeed too early to come to this conclusion. However, we think it is

worthwhile to further study the relation between balancing and moral outrage. In these studies, the role of associative strength should also be considered. The concept of associative strength was introduced by Quinn and Markovits (1998) and applied to defeaters by De Neys et al. (2003b). According to De Neys and colleagues, a defeater has a highly associative strength if it is represented in memory as a good reason to prevent q although p is true. This could be also applied to balancing. A participant will probably only defeat a fundamental right A by another fundamental right B, if B is highly associated in one's memory as a reason to prevent A. For instance, one would defeat right to privacy by right to personal security, if personal security is represented in one's memory as more important than the right to privacy. The only problem with this approach is, however, that it is not clear whether associative strength captures all the case-specific details necessary for balancing. As already described, according to law theory, cases with the same conflicting fundamental rights do not have to end up with the same verdicts. Whether these case-specific circumstances – that are decisive for balancing – are represented in our memory to influence their associative strength requires further investigation. Maybe the associative strength provides some general, case independent, overall weight to balancing, whereas moral outrage is responsible to tune the specific weights in accordance to the case-specific details.

Finally, our studies also show that paradigms from cognitive psychology are useful to investigate questions from other fields. Conditional inference tasks were originally introduced to test people's capacity to reason according to classical logic. In our study, however, we showed that inference tasks are also useful to test accounts from legal theory. We think that this is true for many other areas as well. For instance, inference tasks can also be helpful to study moral reasoning, where – similar to balancing – people also have often to decide between two conflicting principles (e.g., telling the truth or lying to not hurt someone's feelings).

Acknowledgments

This work was supported by DFG grants KN 465/10-1 and KN 465/10-2 to Markus Knauff within the DFG Priority Program "New Frameworks of Rationality".

References

- Alexy, R. (2003). Die Gewichtsformel. In v. J. Jickeli, P. Kreutz, & D. Reuter (Eds.), *Gedächtnisschrift für Jürgen Sonnenschein* (pp.69-78). Berlin: De Gruyter Recht.
- 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. (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.
- Bonnefon, J.-F. (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review, 116*, 888-907.
- 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.
- 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.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23*, 646-658.
- Darley J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior, 24*, 659-683.
- 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.
- Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford, England: Oxford University Press.
- Gazzo Castañeda, L. E., & Knauff, M. (2016). Defeasible reasoning with legal conditionals. *Memory & Cognition, 44*, 499-517.
- Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition, 35*, 2060-2074.
- Habermas, J. (1992). *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaat*. Frankfurt a. M: Suhrkamp.
- Lerner, M. J. (1970). The desire for justice and reactions to victims. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior*. New York: Academic Press.
- Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning, 1*, 121-152.
- Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning, 19*, 346-379.
- Prakken, H., & Sartor, G. (2004). The three faces of defeasibility in law. *Ratio Juris, 17*, 118-39.
- Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68*, B93-B101.
- Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as reasoning in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science, 29*, 919-960.
- 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.