

Testing Descriptive or Prescriptive Conditionals and Differential Effects of Frequency Information

Momme von Sydow (Momme.von-Sydow@bio.uni-goettingen.de)

Department of Psychology, Georg-August-Universität Göttingen, Abt. 1, Goßlerstr. 14
D-37073 Göttingen, Germany

Abstract

The author proposes and investigates different strategies for testing descriptive and prescriptive conditionals in Wason's selection tasks. Two experiments examined the influence of frequency information on the postulated Bayesian and deontic test strategies. The rules in both the descriptive and the prescriptive case were social rules. The results of Experiment 1 showed only some of the predicted differences, and necessitated an extension of the advocated dichotomy. Based on this modification, Experiment 2 completely supported the predicted two types of test strategies. The discussion suggests a classification of different frequency effects that may be rational even for testing deontic rules.

Keywords: hypothesis testing, Wason Selection Task, deontic reasoning, Bayesian logic

Introduction

The philosophical distinction of 'is' and 'ought' is even older than Aristotelian logic and although philosophers agree to disagree about the relation of these two realms it remained fundamental throughout history. It is argued here that this distinction may fruitfully be applied to the psychology of information selection and, particularly, to the intensively investigated selection task, which has become notorious for raising fundamental doubts over human rationality.

The *Wason Selection Task* (WST; Wason, 1966) is concerned with information-selection, particularly in testing the hypothesis of a conditional of the form of "if p then q ". Participants are requested to test a conditional (the theoretical world), for instance "if Ken eats haddock then he drinks gin", against an empirical world made up of four cards. Each card represents a case and, in this example, one side of a card represents some main dish (corresponding either to p or $non-p$) and the other side represents a drink (either q or $non-q$). The four visible card-sides stand for the four logical classes, p ('haddock'), $non-p$ ('salmon'), q ('gin'), and $non-q$ ('vodka'). The participants' task is to select the card(s) that they need to turn over in order to test the rule.

In psychology, the traditional yardstick against which the performance in the WST is measured has been the logical-falsificationist norm of testing hypotheses. Accordingly, the tested conditional has normally been interpreted as a deterministic relation in the sense of an implication: $\forall x p(x) \rightarrow q(x)$. Independent of its content it is demanded that the two potentially falsifying cards, p and $non-q$, are always to be selected (here 'haddock' & 'vodka').

In four decades of intensive research on the WST, particularly two anomalies of the falsificationist-logicist

research program have inspired the rise of a variety of domain-general and domain-specific psychological theories. The first anomaly has been that participants almost never acted in accordance with the falsificationist norm, but generally seemed to select confirming evidence instead. Already Johnson-Laird and Wason (1970) showed that, in abstract selection tasks, 96% of their participants selected a wrong pattern of cards, many choosing a p -and- q -pattern instead. The second anomaly is linked to content effects, the phenomenon that only particular contents 'improves' the performance. In the test of the haddock-gin rule there was no enhanced performance, despite its realistic content. However, an improvement was found for the rule "If a person is drinking beer, then the person must be of full age".

Mental model theory (MM theory) explained both anomalies in a domain-general way, by either complete or incomplete representations of the conditional or biconditional (Johnson-Laird & Byrne, 1991, 2002). Later a *Bayesian approach* of rational hypothesis testing (Oaksford & Chater, 1994, 2003, von Sydow, 2006; see below) proposed an alternative explanation for the first anomaly of the falsificationist norm, the confirmation bias. The second anomaly led to the formulation of two further domain specific theories concerned with obligations or social contracts. *Pragmatic reasoning schema theory* (PRS theory) argued that testing a conditional is not linked to a system of rational reasoning, like formal logic, but to domain specific schemata (e.g., Cheng & Holyoak, 1985). They investigated an obligation schema and a permission schema and found that the former 'facilitates' logical p and $non-q$ answers but the latter elicits clear-cut reversed patterns ($non-p$ and q). *Social contract theory* (SC theory) similarly showed that in social contracts with a cheater detection goal either clear-cut p & $non-q$ or $non-p$ & q patterns could be found, depending on the perspective of the tester (Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992, cf. Fiddick, 2004).

Testing Descriptive Hypotheses and Checking Prescriptive Rules

In the WST debate, Manktelow and Over (1991) first noted a normative difference of testing prescriptive and descriptive conditionals, since only the former cannot be falsified. Nonetheless, previous research has not addressed the normative differences between testing prescriptive and descriptive rules. In a context in which frequency information is varied, it will here be tested for the first time whether tests of descriptive or prescriptive conditionals lead to different typical test strategies. (This idea was suggested earlier in

von Sydow, Metzner, Hagmayer, & Waldmann, 2005, cf. von Sydow & Hagmayer, 2006.)

The distinction of ‘is’ and ‘ought’ has been fundamental to most cultures and religions, and throughout history it has been mirrored by basic philosophical dichotomies such as ontology and ethics, or theoretical and practical philosophy. Descriptive rules describe states of the world (i.e., facts) and can therefore be true or false. In contrast, prescriptive rules state what *should* be done or omitted. They often state what is right or wrong. Representations about what is true or false (traditionally truth tables) have to be distinguished from representations about what is right or wrong (for ‘ought tables’, cf. von Sydow et al., 2006).

It is advocated here that these two types of rules, tested in WSTs, do not only have different meanings, but that they are normally tested by using different *test strategies*. Although content effects in the WST have been observed particularly for prescriptive rules, only few researchers have recognized the importance of rational differences between testing prescriptive and descriptive rules (without testing it directly Oaksford & Chater, 1994, made a similar distinction; cf. v. Sydow, 2006). It is argued here that descriptive conditionals should be tested according to Bayesian norms of achieving maximal information gain, which determines the informativity of certain cells of a truth table mainly *a posteriori*, whereas the testing of prescriptive rules is typically concerned with an *a priori* focus on specific cells of an ‘ought table’.

The *Bayesian norm of testing descriptive rules* (Oaksford & Chater, 1994, 2003; von Sydow, 2004, 2006, cf. e.g., Evans & Over, 2004) takes not only the logical form into consideration but frequency information as well. In Table 1 the basic model of a dependence hypothesis (the validity of the implication) and an independence hypothesis (independence of p and q) is formulated based on fixed resulting probabilities of $P(p_{res})$ and $P(q_{res})$.

Table 1: Dependence (a) and Independence (b) Model of a Conditional with Fixed Marginal Probabilities (v. Sydow, 2002, cf. 2006; Hattori, 2002, Oaksford & Chater, 2003)

a) M_D	q	$non-q$	Marg.	b) M_I	q	$non-q$	Marg.
p	p	0	p_{res}	p	p	p $(1 - q)$	p_{res}
$non-p$	$q - p$	$1 - q$	$1 - p_{res}$	$non-p$	$(1 - p)$	$(1 - p)$ $(1 - q)$	$1 - p_{res}$
Marg.	q_{res}	$(1 - q_{res})$	1	Marg.	q_{res}	$1 - q_{res}$	1

Note: In the cells $p := P(p)$, $q := P(q)$ is used.

If this basic model is combined with further Bayesian calculations (here exactly following Oaksford and Chater, 1994, cf. Sydow, 2006) more p than $non-p$ selections and more q than $non-q$ selections are predicted, as long as one is concerned with standard situations ($P(p) \leq P(q) < 0.5$, $P(H_D) = 1 - P(H_I) = 0.5$). This corresponds to the first anomaly of frequent confirmatory p and q findings in descriptive WSTs if rarity is given. However, from this model

the additional prediction can be derived that for atomic propositions with a high probability ($0.5 < P(p) \leq P(q)$) $non-p$ and $non-q$ selections should increase. Despite the theoretical appeal of this rational Bayesian resolution of the first anomaly, the empirical results have remained problematic (Oberauer, Weidenfeld, & Hörnig, 2004, Oaksford & Wakefield, 2003). Von Sydow (2006) gave an overview of the diverse empirical findings and argued that the problematic results may be due to not establishing the exact preconditions of the model. Based on modifications in the experimental procedure, first support could be provided for different Bayesian models (cf. v. Sydow, 2004).

Whereas descriptive Bayesian models are concerned with the truth or falsity of a conditional, *prescriptive conditionals* are primarily concerned with what ought to be done or to be omitted. The current investigations are based on the theory that the clear-cut logical or illogical selection patterns, p & $non-q$ and $non-p$ & q (second anomaly), shown by PRS and SC theory, could be explained by a more general *flexible deontic logic theory* (FDL theory, v. Sydow et al. 2006, v. Sydow, 2006), which has provided novel findings:

Firstly, the clear-cut selection patterns found earlier for obligation and permission schemas (cf. Manktelow & Over, 1995) were understood based on deontic logic as only *two out of four* possible types of 2×2 ought tables for conditionals with one forbidden cell (in other contexts cf. Beller, 2003, Bucciarelli & Johnson-Laird, 2005). In an experiment, all the predicted p & q , p & $non-q$, $non-p$ & q , $non-p$ & $non-q$ selection patterns were found, corroborating that the (conditional) permission and obligation schemas are only special cases of a more general system of schemas corresponding to traditional deontic logic, including conditional prohibitions.

Secondly, this more complete deontic logic was additionally combined with the variation of the goals when checking the rule (cf. Sperber & Girotto, 2002), eliciting different foci on cheater or cooperator cases of different ought tables (cf. Table 2 for a conditional obligation)

Table 2: Prescriptive Conditional Obligation Rule “If One is a Bachelor One Must Abduct a Virgin” with a Cooperator (Full Circle) or a Cheater Focus (Dotted Circle).

	Does abduct virgin (q)	Does not abduct virgin ($non-q$)
Bachelor (p)	Allowed	Forbidden
Husband ($non-p$)	Allowed	Allowed

Von Sydow and colleagues for the first time showed rational interactions of different deontic rules *and* different goals. Moreover, a double focus condition showed a significant increase of predicted p , q , and $non-q$ patterns. For deontic rules this pattern has not been observed before and comparisons with other patterns made clear that a cooperator focus differs from a mere absence of a cheater focus. It was argued that the focus effects on different ought tables cannot be explained by SC theory, PRS theory or

MM theory – without adopting a deontic logic combined with a flexible focus as postulated by FDL theory.

In this paper, we are concerned with another prediction of FDL theory, which is only linked to the mentioned focus effects. The typical task of a tester of a prescriptive rule – as normally formulated in the WST tradition – is to search, for instance, for cheaters only. In this sense, we are quite literally concerned with a ‘selection task’. The standard instruction of testing prescriptive rules has been to ‘select all cheaters’ (particularly in the tasks of Cosmides & Tooby, 1992). A construction of an analogous task for the testing of descriptive rules would result in the rather unusual instruction to ‘collect all cases where the conjunction p and $non-q$ occurred’ (cf. Wason, 1968; Sperber & Girotto, 2002). Although the ability to focus on different cells of a truth table or an ought table may well be based on a general rational process (Sperber & Girotto, 2002), and there may be mechanisms common to the testing of both kinds of rules, there are differences in how descriptive and prescriptive rules are to be tested in standard contexts.

Here it is argued that given an a priori focus for instance on cheaters in a prescriptive task there should not be the frequency effects as predicted for standard descriptive tasks. For standard prescriptive tasks with an a priori focus, FDL theory predicts not the same frequency effects as it is predicted for descriptive tasks. The postulated difference of these two task types, will here be investigated for the first time with respect to potential frequency effects. However, please note, a complete absence thereof is not necessary in prescriptive tasks. Firstly, from a Bayesian perspective, *within-focus* frequency effects would be rational also for prescriptive rules with a *a priori* focus. But they should only affect cards that refer to the focused forbidden case of a conditional obligation (p versus $non-q$; cf. Kirby, 1994). Secondly, frequency information should be irrelevant for standard prescriptive WSTs, but there may be tasks in which probability determines which cells are to be focused on in the first place (*goal frequency effects* and *sorting frequency effects*). But these probability effects should not lead to the p vs. $non-p$ -effects predicted for descriptive tasks (cf. General Discussion). If such effects are excluded, frequency effects should be present in testing standard descriptive rules but absent in testing standard prescriptive rules.

The distinction may, for instance, explain why Cosmides (1989, Exp. 1, 2; cf. Gigerenzer & Hug, 1992, Exp. 1) found a difference between conditions in which an anthropologist tested for the truth or falsity of a social rule and cheater detection conditions in which participants were concerned with detecting cheaters. This difference does not need to point to a cheater detection module, but it may point to rational ways to solve two quite different kinds of tasks.

However, the postulated differential effect of frequency sensitivity on descriptive and prescriptive WSTs has not yet been tested directly and simultaneously. This is carried out in the following two experiments.

Experiment 1

Method

Design and Participants The experiment had a 2 (prescriptive vs. descriptive task) \times 2 (low vs. high probability) between-subjects design. 80 participants of the University of Göttingen (56 % female, 44 % male) finished the task (cf. v. Sydow, 2006).

Materials and Procedure The tasks were carried out in paper and pencil tests on the campus. The instructions were given in German. The prescriptive and the descriptive conditions varied in the beginning and at the end of the instructions, the middle section was identical.

In the prescriptive task conditions, it was stated that the participants should imagine they were members of a council of elders of a tribe, who had a police function, and whose task it was to *punish* those who had violated the law. In the descriptive task conditions, participants should imagine that they were an ethnologist investigating a tribe’s social rules.

The descriptive rule read, “If someone is a bachelor, then he brings fish to the medicine man” and in the descriptive rule only the modal verb ‘must’ was added: “If someone is a bachelor, then he must bring fish to the medicine man.” In the descriptive conditions, the alternative hypothesis was clarified, the marital status and acts of bringing fish appear together in a completely random way (cf. v. Sydow, 2006).

In all conditions, it was explained that there is a two-sided panel for each male member of the tribe. On one side of each panel, it is said whether the person is a bachelor or husband (mentioning the symbols used later). On the other side of the same panel whether the tribe member brought fish or not was registered (symbols). The instruction continued such that the chief of that tribe shows the participant all fronts of all panels (this corresponds to $P(\text{bachelors})$ and $P(\text{husbands})$). After mixing the panels, he then showed all the backs of the cards ($P(\text{fish})$ and $P(\text{non-fish})$).

There was a high and a low frequency condition in both task types. In all conditions, 20 panels were considered, and hence 20 fronts and 20 backs were shown. The probabilities for the antecedent (‘bachelor’) and for the consequent (‘brings fish’) were ‘0.10 \rightarrow 0.15’ in the low probability conditions and ‘0.85 \rightarrow 0.90’ in the high probability conditions. This actively introduced the preconditions of a model with fixed marginals (cf. v. Sydow, 2002, 2004, 2006). It was emphasized that the panels were mixed between the displays of the fronts and the backs, and hence the mapping of the individual sides of the panels is still unknown.

The participants then had to imagine that the chief of the tribe allows them to turn over only two cards separately.

The participants were then required to tick those cards, which they would turn over to fulfil their task. They could tick one card to be turned over out of the first display (‘bachelor’ or ‘husband’) and another out of the second display (‘fish’ or ‘non-fish’).

Results

The results are reported in Table 3. Comparing the high and low probability conditions in *descriptive tasks*, both the predicted significant increase of *non-p* vs. *p* selections was found (Pearson $\chi^2_{(1)} = 9.23, p < .01$) and the increase of *non-q* vs. *q* selections was found (Pearson $\chi^2_{(1)} = 6.46, p < .05$). In contrast, in the *prescriptive tasks* it was confirmed that there are no *p* vs. *non-p* frequency effects in the low vs. high probability condition (exact Fisher test: $p = .70$). However, deviating from the original predictions a frequency effect still occurred with regard to the *q* vs. *non-q* selections (Pearson $\chi^2_{(1)} = 5.01, p < .05$).

Table 3: Experiment 1, $N = 80$,
Card Selections for the Low and High Probability
Conditions in Descriptive and Prescriptive Tasks

Card selected	(a) Descriptive tasks		(b) Prescriptive tasks	
	Low prob. condition	High prob. condition	Low prob. condition	High prob. condition
	n	n	n	n
P	90 % 18	45 % 9	85 % 17	75 % 15
¬P	10 % 2	55 % 11	15 % 3	25 % 5
Q	65 % 13	25 % 5	60 % 12	25 % 5
¬Q	35 % 7	75 % 15	40 % 8	75 % 15
<i>n</i>	20	20	20	20

Note. Percentage and number of participants selecting each card. Predicted answers in darkened cells.

Experiment 2

Based on the comments of the participants of Experiment 1 it was assumed that the remaining *q* vs. *non-q* frequency effect in the prescriptive task conditions might be based on a reinterpretation of the detection goal (goal frequency effect) or based on the trial to identify all cooperators in order to punish the few remaining cheaters later (sorting effect). These effects would make sense of the otherwise useless frequency information. These kinds of frequency effects would produce exactly the *q* versus *non-q* and no *p* versus *non-p* effects, which were found in the prescriptive tasks. In Experiment 2, it was aimed to exclude these effects, by eliciting an a priori focus more clearly and by excluding the possibility of a sorting strategy (cf. General Discussion).

Method

Design and Participants Experiment 2 again had a 2 (prescriptive vs. descriptive rule) \times 2 (low probability vs. high probability) between-subjects design.

Another 80 participants from the University of Göttingen (51 % male, 49 % female) correctly finished the task.

Materials and Procedure In most aspects Experiment 2 replicated Experiment 1. Only the following details differed in order to exclude the above mentioned strategies with an a posteriori focus. Firstly, in the descriptive and prescriptive

conditions the titles “Testing of Hypotheses” and “Punishment of Rule Violators” was added. Secondly, in the prescriptive conditions it was mentioned that a reward for adhering to the law is not intended and the cheater detection goal was emphasized. Thirdly, it was aimed to prevent the strategy of sorting out the cooperators by emphasizing that the selected cheaters should directly be punished.

Results

Table 4: Experiment 2, $N = 80$ (cf. Table 3)

Card selected	(a) Descriptive tasks		(b) Prescriptive tasks	
	Low prob. condition	High prob. condition	Low prob. condition	High prob. condition
P	80 % 16	50 % 10	75 % 15	85 % 17
¬P	20 % 4	50 % 10	25 % 5	15 % 3
Q	75 % 15	35 % 7	25 % 5	20 % 4
¬Q	25 % 5	65 % 13	75 % 15	80 % 16
<i>n</i>	20	20	20	20

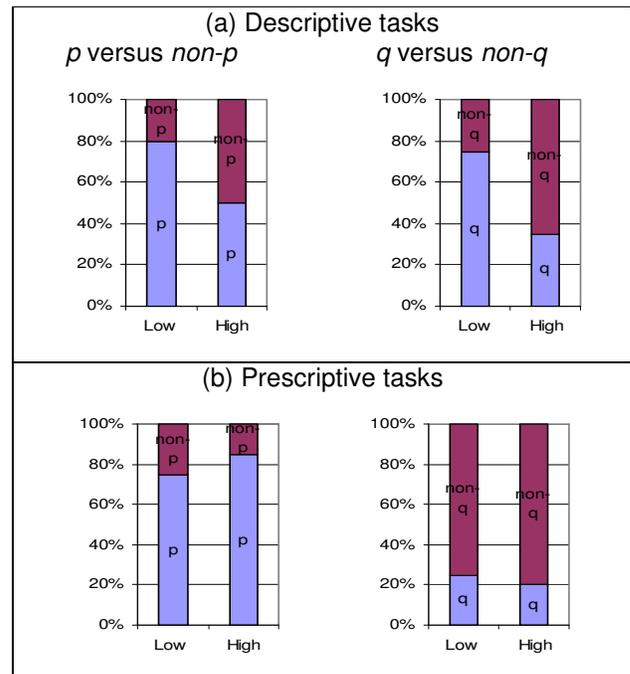


Figure 2. Graphs of the portions of the *p* versus *non-p* selections and *q* versus *non-q* selections in the low and high probability conditions for the two task types.

For the *descriptive tasks* the predicted frequency effects found in Experiment 1 were replicated. In the high versus low probability condition there was an increase in the portion of *non-q* versus *q* selections (Pearson test: $\chi^2_{(1)} = 6.47, p_{\text{one-tailed}} < .01$) and an increase in the portion of *non-p* versus *p* selections (Pearson test: $\chi^2_{(1)} = 3.96, p < .05$). In contrast, for the *prescriptive task* conditions now no difference was found between the low and high probability conditions neither for the *p* versus *non-p* choice (exact Fisher test: $p = .70$) nor for the *q* versus *non-q* choice (Fisher test: $p = 1.00$).

General Discussion

The results of Experiment 2 and also of Experiment 1 corroborate the existence of different test strategies for the standard testing of descriptive and prescriptive conditionals under conditions in which frequency information is provided (cf. v. Sydow et al. 2006, v. Sydow, 2006). I shall argue subsequently that Experiment 1 additionally suggests a more intricate relation between test strategies of descriptive and prescriptive WSTs than originally assumed.

Furthermore, the frequency effects found for the *descriptive* social rules are interesting in their own right. Firstly, the results for the first time show for *social rules* both predicted p versus *non-p* and q versus *non-q* frequency effects. This confirms that the p and q patterns found in the anthropologist conditions of Cosmides (Cosmides 1989; cf. Cosmides & Tooby, 1992) need not be interpreted as irrational but appear to correspond to a Bayesian strategy. Secondly, due to previous problems to achieve these two effects simultaneously, Oxford and Wakefield (2003) argued that these effects seem to occur only when “participants sample the data naturally, i.e. sequentially”, as in their experiment. Oberauer, Weidenfeld, and Hörnig (2004) pointed out a problematic confounding factor in that experiment and themselves provided evidence against the Bayesian predictions. Based on the idea that more positive results may be achieved, if the mathematical preconditions of the models are introduced in detail (e. g. fixed marginals, cf. v. Sydow, 2004, 2006), it was shown here that one may achieve both frequency effects even without sequential tests.

To my knowledge, there are no previous experiments on the effects of frequency information on both descriptive and prescriptive (social) WSTs. Love and Kessler (1995, Exp. 1b) tested both types of rules, but varied only the plausibility of counterexamples, not the probabilities $P(p)$ and $P(q)$. In deontic tasks the number of visible cards was varied by Manktelow, Sutherland and Over (1995) and Kirby (1994, Exp. 4), but both were concerned only with p versus *non-q* selections (within-focus effects) and utility manipulations, not with probability effects.

In Experiments 1 and 2, two different contrasts resulted. In Experiment 1 with a weak a priori focus the p versus *non-p* frequency effect disappeared in the prescriptive conditions, but – in contrast to the descriptive conditions – there was still a q versus *non-q* effect. In Experiment 2, with a clearer a priori focus, both frequency effects disappeared in the prescriptive task conditions.

The results of Experiment 1 were not originally predicted, since formally an a priori focus was used. However, the results could readily be explained by the specific frequency effects for prescriptive WSTs mentioned that are predicted by FDL theory (cf. v. Sydow, 2006, for details): Beside the p versus *non-q* within-focus effects, which cannot be relevant here (due to the limited selection options), there can be rational goal frequency effects and sorting frequency effects also for prescriptive rules.

Firstly, frequency information may have an effect on changing the goal of the task itself. *Ceteris paribus*, the goal

to punish cheaters is plausible if a rule is followed by most people, whereas the goal to reward cooperators may gain plausibility if the rule is normally violated (cf. v. Sydow, et al. 2006, v. Sydow, 2006). In the high frequency condition the minimum possible number of cooperators is here higher than their maximal number in the low frequency condition. Hence, in the low probability condition the goal to reward the few cooperators (q selections) and in the high probability condition the goal to punish the few cheaters (*non-q* selections) may reasonably gain plausibility. Written remarks of some participants confirmed this interpretation. However, this kind of frequency effect for prescriptive WSTs should only be relevant for bachelors (p) leading only to q versus *non-q* effects – as found in Experiment 1.

Secondly, even if one realizes an a priori focus on the forbidden case p & *non-q*, one may try to sort out the few possible cooperator cases in the low probability condition (q selections), exactly in order to efficiently punish the remaining cheaters. Again, p versus *non-p* choices should be unaffected, since independent of whether one tries to sort out cheaters (*non-q*) or cooperators (q), one always aims to check for bachelors (p). This corresponds to the results of Experiment 1. However, sorting effects are only rational if there is a limited universe of discourse and further selections are deemed possible in the future. To exclude this possibility in Experiment 2, it was stressed that no further cards could be selected later on and that one directly aims to punish deviant behavior.

Thus, the contrast between descriptive and prescriptive tasks in Experiment 1 is in line with the claim that different mechanisms were in play in the two kinds of tasks, which is consistent with FDL theory. Moreover, after excluding the discussed factors in Experiment 2, both frequency effects disappeared.

In conclusion, it is wrong to assume that there is only one test strategy for both the standard tests of descriptive and prescriptive rules (cf. v. Sydow, 2004, 2006). With regard to frequency information, there are different selection patterns, which need not be irrational, but which seem to be in line with different rational strategies to test descriptive and prescriptive rules.

Most other theories are either inconsistent with the findings or could explain the results only with substantial extensions. *PRS theory* (Cheng & Holyoak, 1985) inspired the current investigation and would have predicted the p and *non-q* pattern in the prescriptive condition of Experiment 2, but it could explain neither the frequency effects in the descriptive social tasks nor the partial frequency effect in the prescriptive social task of Experiment 1. Likewise, *SC theory* (Cosmides & Tooby, 1992, Gigerenzer & Hug, 1992) cannot explain the frequency effects. Moreover, SC theory claimed that an activation of cheater detection is needed in social contract situations to achieve reliable deviations from normal p versus q selections. However, here novel *non-p* and *non-q* selection patterns were predicted and found. *Relevance theory* (Sperber, et al., 1995; Sperber & Girotto, 2002) also cannot positively account for the reported findings. Although the

formulation of FDL theory owes something to Sperber and Girotto (2002, cf. von Sydow et al., 2006, v. Sydow, 2006), relevance theory, framed in terms of cognitive effect and effort, does not have any immediate predictions for purely probabilistic manipulations. Finally, *MM theory* (Johnson-Laird & Byrne, 1991, 2002) has remained a domain-general theory. Moreover, the explanatory mechanism of MM theory is based on complete versus incomplete representations, not on different test strategies. Particularly, it does not provide any clear predictions for the frequency effects. (According to MM theory there could even be opposed mappings between the theoretically relevant salience of a counterexample and different frequencies, cf. Oaksford & Chater, 1994, v. Sydow, 2006.) Manktelow and Over (1991, 1995) were the first to point out the difference in the result of testing prescriptive and descriptive rules, and I owe much to this distinction. Although the *decision theoretic accounts* of the WST (e.g., Manktelow & Over, 1991, 1995; cf. Evans & Over, 2004) have generally not predicted any particular differences in test strategies for prescriptive or descriptive rules, the investigated distinction is in my view not in principle at odds with an extended decision theoretic approach. Particularly, Oaksford and Chater (1994, cf. 2003) have pre-empted some aspects of the dichotomy investigated here. For prescriptive rules, they advocated a decision theoretic approach with weights for different cells and contrasted this to their Bayesian approach for descriptive tasks. Nevertheless, they have neither explicitly made particular predictions for contrasts between these kinds of rules, nor tested the dissociation supported here.

The current paper contributes to the idea that there are more than one rational test strategies and frequency effects for testing prescriptive and descriptive rules (cf. also v. Sydow, 2004). Here it was shown for the first time with frequency information that standard tests of prescriptive or descriptive social rules may elicit different test strategies and different corresponding selection patterns.

Acknowledgments

I thank Björn Meder and three anonymous reviewers for valuable comments on an earlier draft of this paper. This article draws from a doctoral dissertation (von Sydow, 2006). The work was funded by the Deutsche Forschungsgemeinschaft (DFG, Sy 111/1-1).

References

Beller, S. (2003). The flexible use of deontic mental models. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* (pp. 127-132). Mahwah, NJ: Erlbaum.

Bucciarelli, M., & Johnson-Laird, P. N. (2005). Naïve deontics. *Cognitive Psychology*, 50, 159-193.

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

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163-228). New York: Oxford University Press.

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

Fiddick, L. (2004). Domains of deontic reasoning. *Quarterly Journal of Experimental Psychology*, 57, 447-474.

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

Hattori, M. (2002). A quantitative model of optimal data selection in Wason's selection task. *Quarterly Journal of Experimental Psychology*, 55 A, 1241-1272.

Johnson-Laird, P. N., & Wason, P. C. (1970a). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1, 134-148.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646-678.

Kirby, K. N. (1994a). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition*, 51, 1-28.

Manktelow, K. I., & Over, D. E. (1991). Social role and utilities in reasoning with deontic conditionals. *Cognition*, 43, 183-186.

Manktelow, K. I., & Over, D. E. (1995). Deontic reasoning. In St. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning. Essays in honour of Peter Wason* (pp. 91-114). Hove, UK: Erlbaum.

Manktelow, K. I., Sutherland, E. J., & Over, D. E. (1995). Probabilistic factors in deontic reasoning. *Thinking & Reasoning*, 1, 202-220.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychological Bulletin & Review*, 10, 289-318.

Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: Probabilities do matter. *Memory & Cognition*, 31, 143-154.

Oberauer, K., Weidenfeld, A., & Hörnig, R. (2004). Logical Reasoning and probabilities: A comprehensive test of Oaksford & Chater (2001). *Psychonomic Bulletin & Review*, 11, 521-527.

Sperber, D., Cara, F., & Girotto V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 31-95.

Sperber, D., & Girotto, V. (2002). Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides, and Tooby. *Cognition*, 85, 277-290.

von Sydow, M. (2004). Structural Bayesian Models of Conditionals. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 1411-1416). Mahwah, NJ: Erlbaum.

von Sydow, M. (2006). *Towards a Flexible Bayesian and Deontic Logic of Testing Descriptive and Prescriptive Rules*. Doctoral Dissertation, Universität Göttingen.

von Sydow, M., & Hagemayer, Y. (2006). Deontic Logic and Deontic Goals in the Wason Selection Task. *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (pp. 864-869). Mahwah, NJ: Erlbaum.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135-151). Harmondsworth, Middlesex, UK: Penguin.