Cooperation Detection and Deontic Reasoning in the Wason Selection Task

Momme von Sydow (momme.von-sydow@bio.uni-goettingen.de)
York Hagmayer (york.hagmayer@bio.uni-goettingen.de)
Nicole Metzner (nicole.metzner@uni-goettingen.de)
Michael R. Waldmann (michael.waldmann@bio.uni-goettingen.de)

Institut für Psychologie, Universität Göttingen,
Goßlerstr. 14, D-37073 Göttingen, Germany

Abstract

Domain specific approaches of reasoning have abandoned a universal and normative concept of rationality. In order to account for evidence in the Wason selection task (WST), specifically adapted schemas or Darwinian algorithms have been postulated. We propose a Flexible Deontic Logic Theory, which is a domain specific theory for testing prescriptive rules and nonetheless has aspects of a normative theory. The test of a prescriptive rule does not involve testing its truth, as in the standard WST, but rather checking whether rules are being violated or followed. Unlike descriptive rules, prescriptive rules (e.g., obligations and prohibitions) should be tested on the basis of a deontic logic. These deontic tests involve focusing on cells of an ought table. We claim that this focus is flexible. In contrast to other theories of the WST, we argue that the checking of prescriptive rules is based on the deontic logic combined with a flexible focus either on conforming cases (cooperator detection) or on deviating cases (cheater detection). Our results show that cooperater detection contexts trigger different card selections from cheater detection. These results challenge previous domain-specific and domain-general theories of the WST, but support the view that people flexibly test the validity of deontic rules.

Introduction

Wason Selection Task

The *Wason Selection Task* (WST) (Wason, 1966) is one of the most studied, seminal tasks in the psychology of reasoning. It has been of particular importance also for the rise of domain specific theories of rationality. The WST is concerned with the testing of a hypothesis, typically a conditional in the form of “if \( p \) then (always) \( q \)”. Participants are requested to test the truth or falsity of the conditional in an empirical world of four cards. The visible front sides of the four cards represent examples for all logical categories mentioned in the conditional: \( p, \) non-\( p, \) \( q, \) non-\( q \) (cf. Figure 1). On one side of each card is a \( p \) or a non-\( p, \) on the other side a \( q \) or a non-\( q \). Participants have to select the cards they would turn over in order to test the truth or falsity of the rule.

![Figure 1: The four cards of Wason’s (1966) original selection task with the hypothesis “if there is a vowel on one side there is an even number on the other side”](image)

The traditional yardstick to evaluate answers in the WSTs has been a falsificationist norm (Popper) of correct logical hypothesis testing. According to this universal norm participants should always select exactly the \( p \)- and the non-\( q \)-card, since only they may lead to a falsification of the conditional (cf. the truth table of a conditional, Table 1).

<table>
<thead>
<tr>
<th>( p \rightarrow q )</th>
<th>( q )</th>
<th>non-( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>Non-( p )</td>
<td>True</td>
<td>True</td>
</tr>
</tbody>
</table>

However, from early on empirical results have shown that most participants do not act according to this norm. For example, Johnson-Laird and Wason (1970) found that 96% of the participants in their WSTs selected a wrong pattern of cards, many choosing the \( p \)-and-\( q \)-cards instead.

Domain specific theories, such as the Mental Model Theory (Johnson-Laird & Byrne, 1991, 2002), have tried to maintain the logical-falsificationist norm. They explain the deviations by incomplete representations of the given rule.

More recently Bayesian approaches have tried to explain the predominance of \( p \)- and \( q \)-cards in descriptive WSTs. These selections are shown to be rational, provided that \( p \) and \( q \) are rare (cf. e.g. Oaksford & Chater, 2003; v. Sydow, 2004).

Domain Specific Theories and Cheater Detection

Another class of phenomena, by which traditional domain-general theories have been challenged, are so-called content effects. Thematic rules, such as “If I eat haddock I drink gin,” did not enhance the performance, while social rules, such as “If a person is drinking beer, then the person must be of full age”, did.

According to Cheng and Holyoak’s (1985; cf. Holyoak & Cheng, 1995) *Pragmatic Reasoning Schema Theory*, content effects are due to specific reasoning schemas, which are linked to goals and which are based on abstractions of recurring experiences in the society. Schemata do not always enhance the selection of logical patterns; they may also trigger illogical ones. In the deontic realm they proposed a permission schema and an obligation schema.

Cosmides’ domain-specific *Social Contract Theory* was even more pronounced in abandoning any normative logical basis for reasoning in the WST (Cosmides, 1989; Cosmides & Tooby, 1992; Fiddick, Cosmides, & Tooby, 2000). Instead Cosmides based her theory on evolutionary consi-
The distinction of ‘is’ and ‘ought’ is fundamental. It is central for moral, social and religious regulation systems in our societies. In philosophy, for example, the distinction of ‘is’ and ‘ought’ is older than Aristotelian formal logics and throughout history it is mirrored by the basic philosophical dichotomy of ontology and ethics, or of theoretical and practical philosophy. Despite all differences of philosophical schools the distinction of a prescriptive and a descriptive realm is part of a *philosophia perennis*: it has been made by almost all great philosophers, such as Plato, Aristotle, Augustine, Hume, Kant and by modern analytic philosophers (e.g., Moore, von Wright).

Because descriptive and prescriptive rules are fundamentally different, they need to be tested differently. 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 state what is right or wrong. Hence, the testing of descriptive rules is concerned with the truth or falsity of a rule, whereas the testing of prescriptive rules is concerned with the violation or compliance with a social rule or regulation. Three aspects of this dichotomy can be distinguished:

First, prescriptive rules cannot directly be falsified by conflicting evidence. In the WST context this has first been pointed out by Manktelow and Over (1991). If many people violate the law ‘If one is drunken, it is forbidden to drive a car on public streets’, it is still illegal to do so. Secondly, what is true is not necessarily right, what is false not necessarily wrong. We are hence going to introduce deontic ought tables, complementing descriptive truth tables. Ought tables specify what is right or wrong, allowed or forbidden. Thus the truth of a descriptive rule has no (direct) implications for a similar prescriptive proposition. Thirdly, we advocate different norms for testing these two types of rules. Descriptive conditionals should be tested according to the norms of Bayesian reasoning (e.g.: Oaksford & Chater, 2003; v. Sydow, 2004). Besides the logical form also frequency information needs to be taken into account, to optimize the expected information gain. This potentially makes use of all cells of the truth table (v. Sydow, 2004). In contrast, the testing of prescriptive rules is not typically concerned with all cells of a truth table; instead one should focus on specific cells of an ought table. The typical task of a tester of a prescriptive rule is to check who either acts in accordance or in discordance with that rule (see below).

**Testing Prescriptive Conditionals by Deontic Logic**

We advocate that domain-specific content effects, which are the basis for the illogical domain specific approaches of the WST, could partly be explained and systematized on the basis of a Deontic Logic of Domain Effects. Deontic logic has for long time analyzed logical relations between prescriptive propositions, such as obligations and permissions (see e.g. Hilpinen, 1981). Building on the insights of deontic logic, we propose that the logical structure of a prescriptive rule is represented by an ought table. Table 2 shows a conditional obligation and Table 3 a conditional prohibition. In this paper we will focus on two-valued ought tables. The cells of the tables represent states of affairs or actions, which can either be right or wrong.
allowed or forbidden. A *universal obligation,* for example, like “Thou shalt love thy neighbor as thyself” implies that it is always right to love one’s neighbor and always wrong to hate him/her.

A *conditional obligation* “if p then one must do q” asserts that if it is wrong if p & non-q happens. Assume that a tribal rule says “if you are a bachelor, you must bring fish to the medicine man,” then it is forbidden to be a bachelor and not to bring fish to the medicine man (cf. Table 2).

A *conditional prohibition,* such as “if you are a bachelor, you must not go to the bath house”, forbids that one is a bachelor (p) and one goes to the bath house (q) (cf. Table 3).

Table 2: Ought table of a conditional obligation

<table>
<thead>
<tr>
<th>Conditional Obligation</th>
<th>brings fish (q)</th>
<th>does not bring fish (¬q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bachelor (p)</td>
<td>allowed</td>
<td>forbidden</td>
</tr>
<tr>
<td>husband (non-p)</td>
<td>allowed</td>
<td>allowed</td>
</tr>
</tbody>
</table>

Table 3: Ought table of a conditional prohibition

<table>
<thead>
<tr>
<th>Conditional Prohibition</th>
<th>goes to the bath house (q)</th>
<th>does not go to the bath house (¬q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bachelor (p)</td>
<td>forbidden</td>
<td>allowed</td>
</tr>
<tr>
<td>husband (non-p)</td>
<td>allowed</td>
<td>allowed</td>
</tr>
</tbody>
</table>

Most research on deontic WSTs has been concerned with conditional obligations and conditional permissions and a biconditional understanding of the conditional (cf. Manktelow & Over, 1991; Johnson-Laird & Byrne, 1992; Holyoak & Cheng, 1995; Beller & Sparda, 1998). In contrast we will here be concerned with conditional obligations and prohibitions. Although all mentioned rules are conditionals, they in our view forbid different cases in an ought table, as can be derived from deontic logics. If the task is to look for violations of the rule these different cases provide the normatively correct answers.

Compared to current theories of the WST, this deontic logic approach combines two apparently opposed ideas. On the one hand it preserves the concept of a domain-general logical core (also emphasized by Mental Model Theory). On the other hand it follows Pragmatic Reasoning theory in assuming that different schemata of prescriptive conditionals do elicit different ought tables. However, unlike Pragmatic Reasoning Schema Theory, our theory is based on ought tables that are part of a systematic deontic logic.

**Pragmatic Cell Focus on Cheater or Cooperator Detection**

The second component of our theory postulates that in different pragmatic contexts people will focus on different cases (cf. similarly Liberman & Klar, 1996; Sperber, 2003). The so called ‘cheater detection algorithm’ is, in our view, nothing but a specific focus on the forbidden cells of an ought table. Also perspective effects (Gigerenzer & Hug, 1992) shifted the focus, but only between different the cheater cases (cf. Johnson-Laird & Byrne, 1992; Almor & Sloman, 2000). In contrast, we advocate a flexible focus not only on cheater cases, but also on other cells of an ought table. In deontic contexts, checking prescriptive rules typically involves searching either for individuals who have violated the rule or individuals who have complied with the rule. For example, if generosity is a prescribed norm of a society, then stingy persons should be punished and generous people should be rewarded. Thus, depending on whether punishments or rewards constitute the current pragmatic goal, different cases should be searched.

The concept that the cell focus of testing prescriptive rules is flexible and goal dependent is here combined with Deontic logics. For a conditional obligation the cheater detection focus is on p (bachelor) and non-q (brings no fish to the medicine man), the cooperation detection focus on p and q (brings fish). In contrast for the conditional prohibition rule the cheater focus is on p and q (goes to the bathhouse) and the cooperator focus on p and non-q (does not go to the bath house). This prediction is inconsistent with both, existing domain-specific approaches and domain general approaches. It is particularly critical for theories that postulate innate ‘cheater-detection’ modules (Cosmides, 1989; Gigerenzer & Hug, 1992).

Table 4: Cheater focus in a conditional obligation

The circle indicates the focused cell

<table>
<thead>
<tr>
<th>Conditional Obligation</th>
<th>brings fish (q)</th>
<th>does not bring fish (¬q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bachelor (p)</td>
<td>allowed</td>
<td>forbidden</td>
</tr>
<tr>
<td>husband (non-p)</td>
<td>allowed</td>
<td>allowed</td>
</tr>
</tbody>
</table>

Table 5: Cooperator focus in a conditional obligation.

The circle indicates the focused cell

<table>
<thead>
<tr>
<th>Conditional Obligation</th>
<th>brings fish (q)</th>
<th>does not bring fish (¬q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bachelor (p)</td>
<td>allowed</td>
<td>forbidden</td>
</tr>
<tr>
<td>husband (non-p)</td>
<td>allowed</td>
<td>allowed</td>
</tr>
</tbody>
</table>

**Experiment**

In this experiment, we aim at testing the central prediction of our theory that the selection patterns of checking a prescriptive conditional are determined systematically both by the type of conditional (deontic logic) and by the cheater or cooperator focus (flexible focus). We expected that in a context of possible punishment of rule violators, participants would focus on forbidden cases. In contrast, we expected that a context of rewarding will elicit a search for rule followers and a corresponding cooperator focus.

**Method**

**Design and Participants** The experiment had a 2 (obligation vs. prohibition rule) × 2 (cheater vs. cooperator focus) between-subjects design. Eighty students from the University of Göttingen participated in the experiment, and were randomly assigned to the four conditions.
Materials and Procedure In all conditions we used a deontic Wason selection task. Participants were asked to imagine being a member of a council of elders, whose task was to check whether members of a tribe either have violated or have followed the laws of the tribe. In the two obligation rule conditions the council of elders had to check the rule: “If someone is bachelor, then he must abduct a virgin from a hostile dangerous tribe.” In prohibition rule conditions the following rule had to be tested: “If someone is bachelor, then he is forbidden from fleeing from a battle, which is about to be lost.” Both rules were novel and unfamiliar to ensure that no prior experience with the rule was available. The goals which, according to our theory, should influence the focus on different cells were manipulated by assigning different responsibilities to the council of elders. In the checker conditions of both rules participants were told that “The council of elders is responsible for law enforcement. The task of the council is to punish those who have violated the laws of the tribe.” In cooperation conditions participants were instructed that “The council of elders each year decorates members of the tribe with honor feathers. The task of the council is to honor those who followed the laws of the tribe.”

In all conditions four male members of the tribe were presented to participants for possible checks. These four tribesmen were represented by four cards. Participants were instructed that on one side of each card information was given about whether the man was a bachelor or not, and on the other side whether he has abducted a woman (or has fled from a battle) or not. The cards read: “Bachelor” (p), “Husband” (¬p), “Virgin abducted” (q), and “No virgin abducted” (¬q). Participants had to decide which card(s) they needed to turn over to test whether the rule had been followed or had been violated. They were requested to indicate all cards necessary to fulfill the given task.

Results

Table 6 shows the number of participants who selected particular card patterns. The predicted answers are darkened. For each relevant card selection pattern we separately carried out a hierarchical loglinear analysis with three factors: card pattern (2) × rule type (2) × goal (2).

For the p & non-q-pattern backward elimination (p<.05) retained the predicted second order interaction term ‘card × rule × goal’. The corresponding interaction level (k=3) was highly significant (Pearson $\chi^2(1)=13.08, p<0.001$). As expected there was no significant first order interaction level (k=2, $\chi^2(3)=1.32, p=.72$). A significant main effect level (k=1, $\chi^2(3)=16.4, p<0.001$) only indicated a predominance of other card selections than p & non-q over all conditions. A test of the single parameters of the saturated model showed also that two parameters significantly differed from zero: the second order interaction term ($z=-3.18, p<0.01$) and the main effect term of the card pattern ($z=4.05, p<0.001$).

For the p & q-pattern backward elimination also retained the saturated model. The analysis of each level showed that only the predicted highest interaction level ‘card × rule × goal’ (k=3, $\chi^2(1)=16.47, p<.0001$) was significant. There were no first order interactions (k=2, $\chi^2(3)=3.08, p=.37$) and main effects (k=1, $\chi^2(3)=1.85, p=.60$). The testing of each parameter of the saturated model showed that only the predicted second order interaction (z=3.52, p<.001) and a first order interaction between p&q-pattern and rule type (z=2.27, p<.05) reached significance. The interpretation of the latter, unpredicted term is problematic, because the corresponding level (k=2) was not significant. However, it seems to refer to more p&q-selections in the prohibition than in the obligation rule. This may be due to the phrasing of the prohibition. Terms such as ‘forbids’ might be less open for a cooperator focus than terms such as ‘must’, which was used to phrase the obligation rule.

Backward elimination of the collapsed remaining selection patterns led only to the predicted main effect of a (low) number of this patterns. Only this effect level (k=1, $\chi^2(3)=8.10, p=.04$) and the corresponding pattern parameter (z=2.90, p<.01) was significant. As expected, there was no effect of conditions (k=2, $\chi^2(3)=2.72, p=.44$) and their interaction (k=3, $\chi^2(1)=.58, p=.45$).

Table 7: Number and percentage of participants making each card selection. (Predicted answers in darkened cells)

<table>
<thead>
<tr>
<th>Card selected</th>
<th>Obligation rule</th>
<th>Prohibition rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>P, ¬Q</td>
<td>10 7</td>
<td>1 7</td>
</tr>
<tr>
<td>P, Q</td>
<td>1 14</td>
<td>7 7</td>
</tr>
<tr>
<td>P</td>
<td>2 4 1</td>
<td>0 0</td>
</tr>
<tr>
<td>¬Q</td>
<td>1 0 0</td>
<td>3 0</td>
</tr>
<tr>
<td>P, Q, ¬Q</td>
<td>0 2 1</td>
<td>1 1</td>
</tr>
<tr>
<td>P, ¬P</td>
<td>0 1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>P, ¬P, Q</td>
<td>1 1 0</td>
<td>0 0</td>
</tr>
<tr>
<td>¬P, Q, ¬Q</td>
<td>1 0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>Q</td>
<td>0 0 0</td>
<td>1 1</td>
</tr>
<tr>
<td>¬P</td>
<td>1 0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>P, ¬P, ¬Q</td>
<td>1 0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>¬P, ¬Q</td>
<td>1 0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>¬P, Q</td>
<td>1 0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>P, ¬P, Q, ¬Q</td>
<td>0 0 1</td>
<td>0 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Card selected</th>
<th>Obligation rule</th>
<th>Prohibition rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>15 75%</td>
<td>20 80%</td>
</tr>
<tr>
<td>¬P</td>
<td>6 2%</td>
<td>2 1%</td>
</tr>
<tr>
<td>Q</td>
<td>4 14%</td>
<td>16 5%</td>
</tr>
<tr>
<td>¬Q</td>
<td>14 70%</td>
<td>3 4%</td>
</tr>
</tbody>
</table>

N: 20 20 20 20
Table 7 presents the number and percentage of participants who selected specific cards for each condition.

The corresponding hierarchical loglinear analyses showed that the selection of the theoretically decisive cards (q and non-q) depended on the predicted interaction of rule type and cooperator versus cheater focus: In both cases backward elimination retained the predicted interaction terms. For the q-card exclusively the second order interaction level was found (k=3, \( \chi^2 = 14.96, p<.001 \)), corresponding to the only significant parameter of the saturated model (z=3.70, p<.001). Equally the loglinear model for the non-q-card showed a predicted second order interaction (k=3, \( \chi^2 = 16.98, p<.0001 \)) and only this term of the model varied significantly from zero (z=-3.90, p<.001).

Our most interesting result that the selection of q- or non-q cards was strongly dependent on the type of rule and the goal of cheater or cooperator detection is also depicted in Figure 2.

![Figure 2: Percentage of participants selecting q- or non-q-cards in cheater or cooperator conditions.](image)

**Discussion**

The results of the experiment support the proposed Flexible Deontic Logic Theory of the WST. They provide strong evidence for the claimed interaction of different conditionals based on deontic logic and focus effects. Different selection patterns were preferred in a context of cheater detection versus a context of cooperactor detection. It is possible to elicit clear cut selection patterns not only by conveying the goal of cheater detection, but also by the goal of cooperactor detection. We found reversed trends for the conditional obligation rule and the conditional prohibition rule.

The results particularly favor Flexible Deontic Logic Theory over the cheater detection approach of Social Contract Theory. But without modifications also other theories of the WST cannot fully account for our results. We can here only briefly discuss the other theories of the WST. The results show that neither a flexible focus nor deontic logic can account for the data on their own. Flexible Deontic Logic Theory combines these two mechanisms. The interaction of the type of deontic rule and the goal of cheater or cooperactor detection have not explicitly been predicted by any current theory concerned with the WST.

Social Contract Theory (Cosmides, 1989; Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992) has argued that the activation of a cheater detection algorithm is the only way to produce clear cut WST selections in social contracts. Although perspective effects may perhaps also be interpreted as some kind of ‘focus effects’, such focus effects were explicitly only concerned with different cases of cheating (Gigerenzer & Hug, 1992, cf. e.g. Beller & Spada, 1998). In our study, significant focus effects resulted from both cheater detection and from cooperactor detection. In particular in the obligation rule condition the selection patterns were as clear cut when based on cooperactor detection, as when based on cheater detection.

This finding may seem to contradict previous results of Cosmides (1989; Cosmides & Tooby, 1992) on altruist detection. In these studies it was found that an altruist detection instruction did not lead to any clear cut selection patterns. A closer analysis of the chosen tasks reveals that their finding may be due to the phrasing of the instructions used in their scenarios. Altruistic behavior in these scenarios did not clearly correspond to a specific selection of cards; altruist selections were consistent with both p & q, or non-p & q selections.

At any rate, the results of our experiment support our claim that in testing prescriptive rules, the ‘algorithm’ of cheater detection is only one way of focusing on cells of an ought table. Cheater detection seems to be part of a more general strategy of focusing systematically on different cells of an ought table.

Pragmatic Reasoning Theory (PRS-Theory). Although we follow PRS-Theory (Cheng & Holyoak, 1985; Holyoak & Cheng, 1995) in assuming that deontic conditionals are not interpreted as material implications, we do not regard deontic schemata as being illogical. Firstly, we placed the existing schemata of obligation and permission into a systematics of deontic logics. From this we derived the new ‘schema’ of a prohibition rule, not treated by PRS-Theory. Secondly, and more important, the production rules of PRS-Theory do not allow for systematic differences of a cheater or cooperactor detection context.

Also the Mental Model Theory (MM-Theory) of the WST (Johnson-Laird & Byrne, 1991, 1992, 2002) cannot fully account for our findings. MM-Theory claims that the often found (but ‘incorrect’) selections of q-cards in descriptive WSTs are due only to incomplete representations. First, traditional MM-Theory cannot account for the found effects of different types of deontic conditionals. A more promising approach seems to be the deontic mental model theory of syllogistic reasoning, lately proposed by Bucciarelli & Johnson-Laird (2005). This theory, however, has not yet been elaborated to cover the WST. Secondly, and more important, MM-Theory in our view cannot explain the found quite balanced focus effects of cooperator and cheater detection in the different rules without adding a focus mechanism to their representational mechanisms.

In the framework of the Relevance Theory of the WST focus effects have already been discussed earlier (Sperber & Girotto, 2002, 2003). However, relevance theory does not employ any concept of deontic logic and has not combined deontic logic with focusing. Relevance Theory is compatible with our results if it became extended to include deontic logic. But in this case essential explanatory
mechanisms would not be provided by relevance theory itself, but by the underlying pattern of different ought tables.

Also the specific Decision Theoretic Approach of the WST which at least theoretically refers to deontic logic (Manktelow & Over, 1990, 1991, 1992, 1995) would not have predicted our results. Manktelow and Over (1990) showed that in a rule with high benefit for testing conforming cases and low cost for violating cases people still did not test conforming cases. Partly based on these results, Manktelow and Over (1992, 185) came to the general conclusion “that people are more sensitive to costs then to benefits in deontic selection tasks, particularly the serious costs which can result from being cheated […]” Further research is needed to understand the differences between our successful manipulation, which was based on Flexible Deontic Logic Theory and the unsuccessful manipulation based on the direct variation of utilities by Manktelow & Over (1990, 1992).

In conclusion, our results clearly favor Flexible Deontic Logic Theory over Social Contract Theory and they challenge the other current theories of deontic WSTs.

Acknowledgements
We thank Björn Meder for helpful comments.

References


