

The winter of our content effect:

The role of Relevance Theory in designing facilitating effects in Wason Selection Task studies

Maurice Tan

University of Twente, 2008

Abstract

Relevance Theory gives methodological tools for finding the roots of the so called *content effect* in Wason Selection Tasks. This study used these methods to retool classic tasks of competing content effect theories: Social Contract Theory and Pragmatic Reasoning Schemas. No significant results in the order of Relevance Theory's predictions were found, on the contrary: facilitation effects actually led to lower performance. No significant effects in the direction of either Social Contract Theory or Pragmatic Reasoning Schemas was found either, leaving the debate between these two theories intact. Relevance Theory's claims of designing a Wason Selection Task to be yield very high or very low performance were not substantiated in this study, although hints the method's limitations may have been found in the process.

Primary supervisor: W.R. van Joolingen

Secondary supervisor: C. D. Hulshof

The content effect

One of psychology's most researched topics in the past 50 years has to be the puzzling question of how different elements influence performance in the Wason Selection Task (WST). Originally created by Wason (1966), this deductive reasoning task requires subjects to solve a problem in which they have to apply a conditional rule "if P then Q" on a set of four cards containing "P", "Q", "not-P" and "not-Q". According to formal logic, the only correct selection to this logical problem would be those cards that represent the falsifying instance of "P & not-Q".

Ideally, if all humans were purely rational creatures whose brains are designed to apply formal logic in the most rational way possible, this task would lead them to select the two correct cards displaying "P" and "not-Q" representations of the problem at hand. In the case of the original task, these cards would be: E and 7 (see Figure 1). Interestingly, only an abysmally low percentage of subjects succeeds in making the correct selection in these tasks; typically less than 10-20% (Wason, 1966; Wason & Shapiro, 1971; Evans, 1982; Griggs & Cox, 1982).

Conditional rule: "If there is a vowel on one side of the card, there is an even number on the other side of the card"

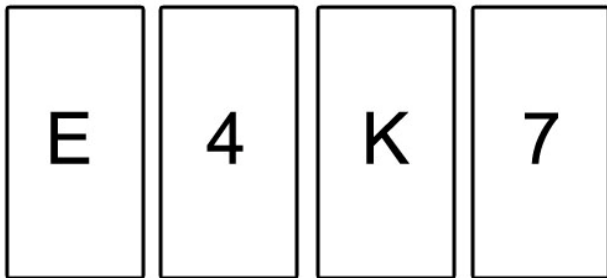


Figure 1: Wason's Selection Task (1966)

At the time Wason created his now famous task, it was already known that tasks that included more meaningful sentences and words led to improved performance when compared to abstract versions of the same task (Wilkins, 1928). This knowledge combined with the results from the WST raised the questions: how exactly does meaningful content lead to improved performance on the WST and exactly what processes make meaningful content *meaningful*?

The quest for answers to these questions resulted in the WST becoming more than just another task in the decades that followed. It was not so much the significance of the task's creation, as it was the implication of its results that shook the foundations of a psychology that focused on syllogistical reasoning in that time (Wason in Evans, 1983). The low percentage of correct performance on the WST starkly contrasted the expectations of rational cognitive reasoning as it was viewed in the 1960's.

It is no surprise then that from the 1970's onwards, studies into the WST tried to come up with explanations for the low performance on an abstract form of the WST while familiarity with the subject appeared to greatly enhance performance. The positive effect of meaningful content on performance

has been called many things. It has been dubbed the *thematic facilitation effect* in certain instances (Griggs & Cox, 1982; Yachanin & Tweney, 1982), but for this study its most common name will be used: the *content effect*.

A lot of explanations have been given for this content effect. Matching bias (Evans & Lynch, 1973; Evans, 1998) tried to explain false P & Q selections through a bias in selecting those options that were given in the task’s conditional rule. Yet the matching bias theory doesn’t hold up well when realistic materials are being used. Mental model theory (Johnson-Laird, 1983) proposes that humans do not usually use formal logic models in everyday problem solving. Formal logic can be used, but it is only one element of mental representations of the problem.

Typically, humans focus on the confirming true states rather than looking for possible falsifications. This is known as confirmation bias (Wason & Johnson-Laird, 1972) or the principle of truth (Johnson-Laird & Byrne, 2002). Whereas these and many other theories focus on a domain-general approach to answering the questions that the content effect raises, two domain-specific theories have tried to explain it in their own way. They are Social Contract Theory (Cosmides, 1989) and Pragmatic Reasoning Schemas (Cheng & Holyoak, 1985)

Pragmatic Reasoning Schemas

Pragmatic Reasoning Schemas (PRS) attempts to explain the content effect through a general set of permission and obligation schemas that govern people’s logical problem solving capability. Instead of consisting of a set of purely syntactic rules, Cheng & Holyoak (1985) describe these schemas as defined in terms of classes of goals (e.g.: taking desirable actions based on any situation) and relationships to these goals (cause and effect or prerequisites for possible actions). The basic permission schema of PRS uses deontic instances of *must* and *may*, while for the most part the permission schema fit the possible P & Q selections as formal logic would (Cheng & Holyoak, 1985). The exception is that PRS tries to block the fallacies of formal logic basic (Table 1), adds deontics to the rules and includes causal schemas as a possibility.

Table 1: Formal rules of inference and rules of the permission schema

For the rule “If P, then Q”:	For action Q and precondition P:
Formal logic	Permission schema
Modus ponens (MP): P, therefore Q	Rule 1: If the action Q is to be taken, then precondition P <i>must</i> be satisfied
Denying the antecedent (DA): not-P, therefore not-Q	Rule 2: If the action Q is <u>not</u> to be taken, then precondition P <i>need not</i> be satisfied
Affirming the consequent (AC): Q, therefore P	Rule 3: If the precondition P is satisfied, then action Q <i>may</i> be taken
Modus Tollens (MT): not-Q, therefore not-P	Rule 4: If the precondition P is <u>not</u> satisfied, then action Q <i>must not</i> be taken

The additions of the deontic elements *must* and *may* add ease of use for everyday problem solving. For instance: Rule 1 can be rewritten as “P is true, therefore Q can happen”, allowing for more context-sensitive (MP) inferences; if you have money to pay for a product, you can buy it. It also prevents the fallacies of (AC) and (DA). For example, Rule 2 states that precondition P is irrelevant if action Q is not taken, blocking the fallacy of (DA): there is no use in denying the antecedent if the antecedent is

irrelevant. Likewise, Rule 3 states that action Q may or may not be taken if precondition P is satisfied, making Q irrelevant and blocking the fallacy of (AC). Finally, failing to satisfy precondition P means that action Q *must not* be taken; if you don't have money to pay for a product, you cannot buy it.

Although the rules of the permission schema, which will be the main PRS focus in this study, share similarities with formal logic, the use of deontics allows for more everyday reasoning than formal logic does. Permission rules are often more useful as heuristics than having to apply formal logic: why spend more computational resources in trying to find the validity of a statement, if you can just compare it analogously to an everyday heuristic and find the answer using much less time and mental resources? Furthermore, causality plays a big role in everyday reasoning (for a review of the influence of causality on reasoning, see Goldvarg & Johnson-Laird, 2001). Events are often perceived as having a single cause, leading to a possible fallacy of seeing an effect as evidence for a possibly invalid cause, an (AC) fallacy. Using permission schemas, people would consider the effect as only one possible result of the cause. A cause might be the reason that an effect is noticed, it also might not be. If you get stomach aches after eating spoiled food, then spoiled food is not necessarily the only cause that created a stomach ache.

Permission schemas are both context sensitive and domain specific. This means that evoking a permission schema in a WST that has thematic content, could lead to facilitation through the use of permission rules. If a subject is familiar with the context and has experience with the specific domain, exposure to such a problem would evoke permission schemas in the subject and lead to increased performance levels, when compared to subjects where permission schemas would not be evoked. Only adding a rationale to the problem would help any subject to evoke a permission schema and perform well on a WST.

One classic example of this is the Stamp Problem (Cheng & Holyoak, 1985), where recent experience with a specific postal rule led a group of students in Hong Kong to performance levels of 90% compared to 60% levels of Michigan students who never heard of the rule. Adding a rationale to the task, giving subjects more information about why they should care about the rule, led to performance levels of 90% for both groups. This meant that context sensitivity was definitely a factor in explaining performance, through experience with the specific context. It also meant that adding thematic content in the form of a rationale facilitated problem solving for those students that did not know about the rule beforehand: contrary to the syntactic view, thematic content influenced performance while the rule remained the same. According to PRS, adding a rationale and keeping context specific prior knowledge in mind can help evoke permission schemas in subjects and therefore lead to increased performance and explaining the content effect.

Among criticisms of Pragmatic Reasoning Schemas is the claim that the used experiments contained a negation in the rationale of the task, leading to more falsifying strategies regardless of the evocation of a permission schema (Noveck & O'Brien, 1996). PRS studies as created by Cheng & Holyoak contained methodological flaws that would be addressed by Cosmides (1989) in her formulation of Social Contract Theory: subjects were allowed to flip back and forth between tasks of the Stamp problem, which could have lead to a transfer effect in the process. In absence of such transfer, no content effect would take place (Griggs & Cox, 1982).

Social Contract Theory

Social Contract Theory (SCT) stems from the school of Evolutionary Psychology and looks to the past for explaining the current-day dilemma. According to the proponents of this theory, the failure of implementing pure logic in solving selection tasks would lie in hardcoded human behavior for reviewing social contracts and more importantly, cheating detection algorithms (Cosmides, 1989). Because prehistoric man was forced to enter social contracts for trading food and other resources in order to function in settlements, it was very important to notice cheating of these social contracts. For example: trading food for clothes in winter could lead to death, when the other party would accept the food but not give anything in return, making survival a matter of being able to detect possible cheating.

Whether or not all aspects of social contracts and cheating detection are remnants of an evolutionary past is still unclear to this day. What is clear however, is that the social contract element has largely been dropped in favor of a focus on cheating detection, which remains to this day one of the more influential explanations for the content effect (Gigerenzer & Hug, 1992; Liberman & Klar, 1996).

Cheating detection is an important concept because it embodies the search for a falsifying instance; overriding confirmation bias in a similar way as prior knowledge of similar realistic content overrides this bias in a familiar context. While still being called SCT, the theory nowadays is all about cheating detection and perspective change. The latter component states that putting a subject in the perspective of someone who can cheat or can be cheated, opposed to the perspective of someone who merely observes actions and does not have any reason to care either way, will lead to high levels of facilitation (Gigerenzer & Hug, 1992). This facilitation remains the same if perspective is switched between versions of the same task while other elements remained equal: the perspective of someone whose job it is to look for cheaters would lead to more falsification strategies and better performance in selecting the correct falsifying instance in a WST rule.

A classic SCT task is the Cassava Root problem (Cosmides, 1989). Subjects are given a thematic story about a fictional tribe on a pacific island that uses Cassava Root as an aphrodisiac. Only married men are allowed to eat the Root, who are distinguishable by the facial tattoos they receive when marrying. In one version, subjects were asked to find out who would be breaking the rule: those who ate Cassava Root but did not have tattoos on their faces. In another version, subjects were told that an anthropologist observed that tattooed tribesmen ate Cassava Root while un-tattooed men did not eat it. Subjects were also given a version of the task that contained a rationale that matched geographical island living locations to forced food diet. They were then asked to select which combination of cards would break the rule "if a man eats cassava root, he must wear a tattoo on his face". The idea behind these tasks were that availability theories of reasoning that claimed that prior experience or knowledge of a task or problem were enough to guarantee facilitation, would have the same hypothesis for either version of the task: subjects would not have experience with the unknown tribe or their customs and therefore performance would be low in either version. SCT meanwhile, would expect high performance in versions that contained social rules of exchange and possible cheating.

A problem with the Cassava Root task, and many other tasks in Cosmides' original 1989 study, was that subjects were so much primed into a falsifier perspective that it was not comparable to tasks of rival theories like PRS. As long as tasks from either PRS or SCT are not comparable, no conclusions can be drawn on which one would be the 'better' theory for explaining the content effect. Regardless, SCT and PRS had an academic dispute on which one was better throughout the 1990's, each trying to discredit or disprove the other through claims of methodological inconsistency and mixing up facilitating elements in their design (for a comprehensive review of methodological flaws, see Liberman & Klar, 1996).

Relevance Theory

Another theory, Relevance Theory (RT) entered the fray in the mid 1990's, attempting to settle the dispute over the content effect in a novel way (Sperber, Cara & Girotto, 1995). Rather than coming up with another single all-explaining concept or theory for the content effect, RT proposes that "relevance" (a general context-specific facilitating effect) can be achieved through two major components: cognitive effect and cognitive effort.

One confusing element of RT is that the meaning of relevance as it is used in language is not the same as the meaning of relevance in Relevance Theory. The closest element of RT's concept of relevance that comes close to matching the semantic definition of the word 'relevance' is *cognitive effect*. "How does this relate to my situation?", "Is it worth investing time in solving such a problem in a real life situation?" these are the questions that subjects tackle when evaluating the cognitive effect of any problem solving activity. As such, any abstract task will always have little cognitive effect and any thematic task that includes a rationale for the rule will always have higher cognitive effect.

This makes the domain-specific context of a task and the inclusion of a fitting rationale key to manipulating the cognitive effect of a WST. In relevance tasks, making P & not-Q selections more salient and accessible in the subject's mind is a way to use cognitive effect to basically 'streamline' the subjects mental problem solving path towards the right selection. Another way to use cognitive effect to the same ends would be to render incorrect selections less salient.

Where cognitive effect is a means to guide subjects to the correct selections by making it easier for them to reach these selections and make them more salient, cognitive effort focuses on mental effort rather than accessibility and salience. Cognitive effort is the amount of mental resources that a subject requires in order make the correct P & not-Q selections in a WST. Basically, every element in the task's rationale, the conditional rule or the selection options themselves that make it hard for someone to process it will lead to high cognitive effort. Anything that makes it easier will lead to less effort and, according to RT, will lead to a facilitating effect on a selection task. If one were to process all available literature on reasoning and the WST in particular of the past 40 years, it would theoretically be possible to make a super facilitating task in terms of cognitive effort. Ironically, such a task already exists in the form of the Drinking Age problem (Griggs & Cox, 1982) where people reach over 90% levels of correct selections time after time.

Relevance Theory goes a long way in turning the content effect from a single effect into a general facilitating effect that has its origins in a number of facilitating factors. Instead of matching a task to fit a certain element like 'cheating detection' or by including a permission schema, RT proposes that the problem solver can be guided to the correct P & not-Q selection using a very large amount of manipulations in any WST. RT does not even oppose P & Q confirmation bias, as long as thinking of P & Q is just one step towards finding the ultimate falsifying selection.

Yet the question remains how and if RT can explain the effects found in both PRS and SCT studies. Or at least those effects of the studies that have not been overly criticized on their methodological shortcomings. One of the major problems in the entire field of WST literature is how most studies create their own method of testing and their own form. New approaches have continuously been tested through the creation of new tasks to disprove studies that use entirely different tasks, or tasks that included different facilitating elements. RT offers the possibility of creating similar tasks with minor variations between versions, while explaining each and every component that is used in the task's design. It also offers the possibility of reviewing classic tasks and analyzing them in terms of RT's cognitive effort and effect.

The current study will attempt to do the latter. Using RT's views on creating more facilitating and less facilitating versions of classic tasks, how will classic PRS and SCT tasks perform when they are adapted to RT task creation methodology? If PRS and SCT respectively are correct in their claims of explaining the content effect in their own way, it shouldn't matter if their classic tasks are made more or less facilitating using RT methods of manipulating cognitive effect and effort that do not interfere with either PRS or SCT predictions. A permission schema should still be facilitating whether it has been redesigned to have less cognitive effect and more cognitive effort. If cheating detection is so important, it should have a large effect on performance regardless of cognitive effect and effort manipulations.

However, if the results deviate from these original tasks and follow RT predictions, it would indicate that PRS and SCT explanations for the content effect are both incorrect and that RT has more explanatory power than either theory. As long as RT manipulations are the essentially the same for both a PRS and a SCT task, it also allows a comparable set of task versions to see which of the two older theories is more facilitating in the same task format. For this study, the prediction will be that both classic tasks will lead to more correct selections when constructed to have more 'relevance' according to RT methodology, than these tasks originally elicited. When they are constructed to have less relevance, they will likely lead to lower performance than the versions which have more relevance, just like Relevance Theory states.

Adding relevance and irrelevance to classic PRS and SCT tasks

The construction of the "relevant" condition of the classic tasks is no easy undertaking. Since RT stresses the importance of guiding the subject through the task in a way that has a maximum effect and a minimal effort, it can be easy to end up with entirely different tasks. Doing so would destroy any possible comparison between the remodeled tasks and the original PRS and SCT tasks. This section will describe the construction of *relevant* and *irrelevant* conditions of two classic tasks that contain the core theoretical essence of the respective theories: permission schemas and cheating detection in a social context. To make both classic tasks fit the RT design of creating salience or obstructing the path to falsification, both tasks will have a narrative style that mirrors the tasks that Sperber, Cara & Girotto used in their original experiments (1995).

Relevance in Relevance Theory

As a template for creating relevance and irrelevance (or the lack of relevance) in a task, the Bachelors problem (Sperber, Cara & Girotto, 1995) was used. This task is a WST at heart and contains two clerks who disagree on a rule. The task's context is a town where volunteers for a school have to fill out cards that the clerks can check. In the relevance condition, one of the clerks claims that men who are bachelors do not like children. Therefore any male volunteers will be married. In the irrelevance condition, the claim is that men with dark hair love children, therefore all male volunteers must have dark hair. Because this effectively creates two different rules as well, only one rule was kept when changing the classic PRS and SCT tasks. If Relevance Theory is correct in its assumptions, then changes in effort and effect alone should lead to significant facilitation.

Pragmatic Reasoning Schemas

Cheng & Holyoak (1985) created an abstract permission task which was later adapted by Noveck & O'Brien (1996) in their critique of Cheng & Holyoak's experiment methodology. Appendix I shows this original task as it was used by Noveck & O'Brien.

Although this task is abstract in form, the use of permission schema structures in the task itself led to performance of 67% correct responses compared to the sub-20% that a fully abstract task would elicit. Apart from the possible elicitation of a permission schema, a 'detective set' of instruction was also included in the task: subjects were asked to specifically find a violation of the rule rather than evaluating it. Such a 'detective set', or the perspective of an authority looking for possible violators, can lead to improved WST performance by itself (Yachanin & Tweney, 1982; Griggs, 1984; Valentine, 1985; Yachanin, 1986). For this study, we will note the existence of such an element in the task, but will keep it intact so that the relevant and irrelevant conditions will remain comparable to the classic task.

Some other elements would have to remain the same as well. Liberman & Klar (1996) give a good discourse of elements that have to be checked in WST methodology. The direction of the task's two versions need to be the same and unidirectional. Subjects have to see the task as "if P, then Q" and not as "if Q, then P" or as biconditional. Perspective must remain the same. The subject has to be an authority figure who checks the rule in both conditions. Falsification must be an equally optimal strategy as other selection options: P & not-Q must not be rendered impossible or unlikely compared to other selections. This last element stems from Kirby (1994), who states that card selection depends on the perceived utility, where utility is defined as the product of the values of the expected outcomes and their estimated probabilities. If the utility of any selection would be higher than P & not-Q selections in everyday life situations, the chances of P & not-Q selections for everyday life content in a WST suffer from it.

In order to make this classic PRS task more and less relevant, the following elements were changed. The probability of P & not-Q being the correct selection will be increased for the relevance condition, while it will be decreased for the irrelevance condition. A deterministic relation (it must absolutely be true that...) will be used for the relevance condition while a probabilistic relation (it's highly likely that...) will be used for the irrelevance condition. A probabilistic relation between P & not-Q should lead to more effort in evaluation the problem (Chapman & Chapman, 1959).

Also, an explicit negation was used in the relevance condition, since it facilitates a task when it also contains a permission schema (Noveck & O'Brien, 1996). The explicit negation can help subjects to go from a mental P & Q selection to a P & not-Q selection, in line with increasing cognitive effect and compatible with RT's concept of guiding the subject through possible confirmation bias towards correct falsification (Sperber, Cara & Girotto, 1995).

The full tasks can be found in Appendix IV.

Social Contract Theory

The Cassava root task (Appendix II) as constructed by Cosmides (1989) has been used in many other studies that tried to explain the results (Gigerenzer & Hug, 1992; Libermann & Klar, 1996). The rule of this task was: "if a man eats Cassava root, he has a tattoo on his face". The social contract element was

that only married men had tattoos on their faces, therefore only married men could eat the root. Performance on the original task was 75% correct (Cosmides, 1989) and 81% correct (Liebermann & Klar, 1996) respectively. This made it a good task to adapt to RT construction: if relevance is no factor at all, the same results could be expected.

Cassava root was changed into Koriko root because cassava is something that Dutch subjects might be familiar with; the goal of the original task was to use only elements that would be alien and unrecognizable to subjects. Many if not most other elements in both versions of the task were kept identical. The perspective was still one of a third party violation detector who had no personal influence in either result, just like in the original task. The root still had aphrodisiacal properties and the rationale for married men being the only group to be allowed to eat them remained the same: allowing unmarried men to eat the root might lead to premarital sex and unplanned pregnancies. The direction of the task was also kept unidirectional; there could be no mistake that eating the root would lead to a man having a tattoo on his face.

To create relevance and irrelevance in the Cassava root task, the relation between tattoos and eating the root was made deterministic in the relevance condition. The relation was made probabilistic in the irrelevance condition by adding a distraction story. In the relevance condition, two anthropologists get into a Bachelor problem style debate about men without tattoos possibly eating Koriko root. In the irrelevance condition, the debate is about men with a moustache eating it. The moustache is part of a fashion element on the island in the task rationale, while having nothing to do with the rule itself. Nor is it ascertainable if a man with a moustache would be married or not, leading to confusion through the alternate possible reasoning, and creating less effect and more effort in this version of the task. Finally, just like in the PRS task, explicit negations were used in the relevance condition while they were absent in the irrelevance condition.

See Appendix V for the full, adjusted tasks.

Drinking age task

The Drinking age problem by Griggs & Cox (1982) is used as a control task for multiple reasons. In short, this task puts the subject in the perspective of a rule violation checker who has to control the rule "If a person is drinking alcohol, he/she is 18 years or older". Firstly, this task should have a near 100% success rate for students, who should all be familiar with the local drinking age. Secondly, having any control task for both conditions is recommended to filter the participating subjects; if they can't do the drinking age task, it raises doubts on the reliability of their response on the other tasks. If subjects fail the drinking age task but have correct selections on the relevance tasks, it would be more likely that they just guessed an answer than that they are logical geniuses who don't know anything about the age limit for alcohol.

See Appendix III for the Drinking age problem and Appendix VI for the Dutch version that was used in this study.

Method

Subjects

32 students and recently graduated students from University of Twente were randomly split into two groups. These students were all from courses other than Psychology who hadn't done a WST before. One group performed the relevance condition of the PRS and SCT tasks, followed by a control Drinking Age task (Griggs & Cox, 1982). The other group performed the irrelevance condition of PRS and SCT tasks, followed by the same control task. Each group was split in half to control for possible transfer between PRS and SCT tasks: one group received the PRS task followed by the SCT task, while the other group received the SCT task first, followed by the PRS task.

Materials

All tasks were provided in Dutch and were put on a website in order for subjects to participate in their natural environment of everyday thinking. The relevance group randomly received either the PRS or SCT tasks first, followed by the other. The last task was always the Drinking Age task. The irrelevance group followed the same method of selection and randomization. Pictures of the cards that were to be selected were ordered differently between groups as well, essentially leading to four different sets of cards for all conditions.

The tasks and pictures were put on an online survey website and led to four sets of tasks, two sets for the relevance and irrelevance conditions divided into two sets each to counter transfer.

Subjects were asked to participate via email and other Instant Messaging methods. When subjects confirmed they wanted to participate, they were sent a link to one of the four sets of tasks. They were explicitly asked to complete the tasks at a certain time during the day, when the subjects were in a normal state of cognitive ability (ie: not at night). Before starting the experiment online, they were also asked to turn off any music and ignore distractions like new e-mails or Instant Messaging programs.

For most students this would be the same situational context as they would normally study in, when working on a computer. If any of the subjects disregarded the request to turn off their music, they would likely have music on during other computer-related work as well, and therefore participate in their normal state of mind. All subjects were expected to follow the requests, and if not, they would probably still perform the tasks in their natural cognitive states rather than an artificially created one.

After completing the survey, the experimenter checked if all of the survey's boxes were filled out and in most cases did a small post-test interview. This was done to gauge the subjects' focus on performing the tasks and to see if they had any specific problem with any of the tasks. These small interviews are not included in the results.

Results

The correct (P & not-Q) and incorrect (other) selections on both conditions of the tasks can be found in Table 2. Relevance and irrelevance conditions did not show significantly better performance for total results between conditions ($\chi^2= 2.64$ (df=1), $p=0.10$). Results for the PRS task did not show any significant difference between the relevance and irrelevance conditions ($\chi^2= 0.05$ (df=1), $p=0.82$) although the results for the SCT task between the two conditions did ($\chi^2= 4.27$ (df=1), $p=0.039$); results for the SCT Irrelevance condition were significantly higher than those in the SCT Relevant condition.

Table 2: Frequencies of correct (P& not-Q) and incorrect (other) selection patterns

	PRS relevant	SCT relevant	PRS irrelevant	SCT irrelevant	Total Relevance	Total Irrelevance	Total PRS	Total SCT
Correct	7	5	6	9	12	15	13	14
Incorrect	11	13	8	5	24	13	19	18
% Correct	39	28	43	64	33	54	41	44

Within the total results for the relevance condition, PRS and SCT did not show any significant difference between each other ($\chi^2=0.5$ (df=1), $p=0.48$). Within the total results for the irrelevance condition, no significant difference was found either ($\chi^2=1.29$ (df=1), $p=0.256$). Between the two conditions, totals for PRS and SCT also did not show any significant difference ($\chi^2=0.06$ (df=1), $p=0.81$). However, a 3D (2x2x2) chi-square analysis of correct and incorrect performance on the relevance/irrelevance conditions of PRS and SCT tasks did yield significance ($\chi^2=4.926$ (df=1), $p=0.026$).

Table 3: Selection performance split into correct (P & not-Q) selections and other incorrect selections

Selection:	PRS relevant	SCT relevant	PRS irrelevant	SCT irrelevant	Total Relevance	Total Irrelevance	Total PRS	Total SCT
P & not-Q (correct solution)	7	5	6	9	12	15	13	14
P & Q	4	6	2	1	10	3	6	7
P & not-P	1	0	2	0	1	2	3	0
Other incorrect selections	6	7	4	4	13	8	10	11
Total Incorrect selections	11	13	8	5	24	13	19	18
Percentage (P & Q) of total incorrect selections	36%	46%	25%	20%	41%	23%	32%	39%

P & Q selections, while not being correct selections for the problem at hand, made up a large part of the incorrect selections as can be seen in Table 3.

The results on the Cassava root task originally displayed 75% correct selections (Cosmides, 1989). When stripped of the cheating aspect, the task only yielded 36% correct selections (Gigerenzer & Hug, 1992). The removal of cheating was done by turning the task rationale into an objective case of two options, eating the root or eating something else, and a rule that governs who eats what. In the Koriko root task used in this study, there was definitely a possibility of cheating. However, the cheating mirrored the method Liberman & Klar used to make the original SCT tasks 'unconfounded' (1996).

By confounding, they mean the process of using different methodological elements that could influence performance, which they criticized previous SCT studies for. The results from their unconfounded Cassava root task were 41% correct selections. The 43.2% correct selections in the total of SCT results are close to the performance found by Liberman & Klar, although it is still far from the high performance levels that the earlier SCT studies had found.

The original abstract permission task as used by Cheng & Holyoak (1985) yielded 61% correct performance, which they replicated with 62% performance levels (1989). The slightly altered version by Noveck & O'Brien (1996) resulted in 67% correct performance. This is a far cry from the 36.3% that was found in the PRS relevance condition.

The Drinking Age task was considered very easy for participants in all groups. Only 2 subjects failed on this task in both relevance and irrelevance groups, leading to 88.9% and 85.8% correct selections for respectively.

Conclusions

The expectation that performance in the relevance condition would be higher than performance in the irrelevance condition was not found in this study. In fact, the opposite was found: the irrelevance condition tasks led to significantly better performance for the Social Contract Theory task, although no significant result was found for the Pragmatic Reasoning Schema task. This indicates that the Relevance Theory approach of modeling tasks to guide the subject to a final selection of P & not-Q cards did not work in this case.

One possibility is that while the relevance and irrelevance task construction elements worked for the tasks like the Bachelor and Virgin mother problems (Sperber, Cara & Girotto, 1995), they do not necessarily work when applying them to existing tasks. The existing tasks already contain a lot of elements that have to be contemplated in the problem solving process, perhaps too many elements. While the RT tasks revolve around a small set of elements that are designed to render the solution salient in the elements themselves (bachelors not being married, virgins not have had sex), changing facilitating or inhibiting elements of Relevance in existing tasks does not work in the same way.

While RT states that cognitive effort and cognitive effect are the key to guiding the subject to the solution, an overabundance of elements to consider in any task could lead to an increase in cognitive effort. At the same time, such an overabundance could lead to clutter and confusion while evaluating the cognitive effect of the elements in question, leading to a decrease in effect. Both would have a detrimental effect on performance.

Another possibility is the incompatibility of different facilitators. Facilitators like negation and directionality were implemented because they can be explained as effort/effect changers in RT terminology. However, the use of both or many facilitating elements in one task does not mean that the possible results will just be the sum of their individual facilitating effects.

Although this could explain the low performance on the relevance condition, it does not explain why correct performance was greater on the irrelevance condition of all tasks. Results for PRS relevance and irrelevance conditions did not change a lot between conditions. Both were well below the expected results in the 61%-66% range for facilitating PRS tasks. Results for the SCT conditions however, changed from 27.8% in the relevance condition to 64.3% for irrelevance, contrary to predictions. The relevance condition even used a “bachelor” rationale in its task description, priming subjects to look for non-married men who are violating the rule by eating Koriko root. There should be no reason why the results would be this much higher on the irrelevance condition. Apart from added facilitators, the only change to the task would be the length of the relevance condition; it is a bit longer because of added elements to counter any thoughts leading away from the P & not-Q cards. It is possible that these added elements made the task harder to understand instead of easier, though not very likely.

RT’s concepts of cognitive effort and effect are built on 30 years of research into facilitators of the content effect. The literature only suggests that a facilitator increases performance on a WST when compared to a version without this facilitator. Combining multiple of these facilitators to create less effort and more effect for the relevance conditions, and more effect and less effect in the irrelevance conditions, should have led to a higher performance on the relevance condition. Not the other way around.

Perhaps from a connectionist viewpoint, increasing the context of a rule-based problem will increase the network and its complexity and therefore lead to less selections of P & not-Q (Leighton & Dawson, 2001). This same viewpoint states that a domain general neural networked approach can explain the effects found in domain specific SCT studies (Eliasmith, 2005; Parsell, 2006). Perhaps the

length of the relevance condition SCT task impaired memory and led to decreased performance on the task. Other than that, there is little guidance in the known literature for the found effect on irrelevance's significant victory over relevance in the case of the SCT task.

Comparing the PRS and SCT results regardless of relevance, neither theory seems to be a clear victor in the debate between the two. No significance was found when comparing the two rival theories. Results for PRS were around 19%-23% lower for the highest percentage of correct selections from the two altered versions, than the original unaltered task it was built on. For SCT, the highest percentage of correct selections (in the irrelevance condition) was only marginally lower than in the original task. It would seem that the debate between the two will not be impacted by the results from the current study.

Of special interest though, are the results of the P & Q selections (Table 3). RT predicts that the selection of P & Q as a mental stepping stone is allowed and even recommended. As subjects are guided towards the correct falsifying selection of P & not-Q cards, they can first make a confirmation bias selection of P & Q cards. But the end result should ideally be that this initial confirmation bias is overruled after further analysis by the subject, leading them to select the falsification instance. The large amount of P & Q errors and the lack of P & not-Q selections could indicate that subjects are guided along the path of Relevance, but do not complete it. The residual P & Q selections were all higher than those found in the original task studies in the case of PRS (Cheng & Holyoak, 1985; 1989; Noveck & O'Brien, 1996) and SCT (Cosmides, 1989; Liberman & Klar, 1992). The highest percentage of P & Q selections in those studies were around the 20% level, while those in this study were at 41.2% on average in the relevance condition and 23% in the irrelevance condition.

This could indicate that the irrelevance condition does indeed mirror the same amount of confirmation bias errors as the original task, while the relevance condition's increase in this type of error may indicate that relevance was almost achieved. If the tasks in the relevance condition were perhaps constructed to have even more effect and require less effort, the excess levels of P & Q selections could be turned into correct P & not-Q selections. Then again, these P & Q selections could all just be a form of matching bias.

Even if some P & Q selections were actually incomplete paths to the correct P & not-Q selection, this would have had no effect on the relationship between relevance and irrelevance in this study. It would only make the difference between totals in the relevance and irrelevance conditions even smaller, and not create more explanatory power for relevance. In the end, no substantial proof was found in favor of Relevance Theory's methodology of effort and effect manipulation. Just adding more elements into a task to create more effect and less cognitive effort does not seem to be the magic solution to the content effect in the WST.

In theory, the real effort it costs to do these tasks could be measured in reaction times. Although cognitive effect would have to remain the same for both cases if the effort is to be measured; cognitive effect could also have an effect on the reaction times. And even then, there is nothing in Relevance Theory that predicts how long a subject would take to process effects and effort viewed as 'relevant' as the theory sees it. Even if all subjects would be interviewed after completing the tasks, it would be hard to make a distinction between cognitive effect and effort from the interviews. If a subject complains about the difficulty, this provides no information about either of the two elements. In the current study, subjects mostly complained about the length of the set of tasks and they did it equally for all versions.

The RT method of dividing facilitators into cognitive effect and cognitive effort is flawed on this front. It is easy to claim that one or the other previously proven facilitator falls under either effect or

effort, but it is much harder to measure which element does what. In RT's view, the two components work together in unison to guide the problem solver to the correct solution, removing obstacles of effort and providing scaffolds of effect along the way. Perhaps the theory is too ambitious for its own good, the amount of variations that may happen implicitly because of interaction between elements of effect and effort, makes it a lot harder to locate why a content effect is found: contracting the reason the theory was created in the first place.

RT's model of facilitating elements does fit better with connectionists models though. Cognitive effect can easily be translated as the strength of connections between relevant nodes or networks, while effort can be seen as the number of epochs it takes for a neural network to reach the correct solution. As such, perhaps future connectionist endeavors may shed some light on the precise influence of effect and effort in the scheme of relevance. Using PDP approaches to testing could remove the need to create a participant group for every possible variation before expectations of the values of effect and effort can be created. It would certainly be helpful to test the impact of all the chosen facilitators before they are grouped into a task and tested. Although it is still unlikely that there would be a prediction that irrelevance conditions lead to higher performance than relevance conditions.

Because no significant difference was found between totals from Pragmatic Reasoning Schemas and Social Contract Theory tasks, it looks like Relevance Theory cannot explain both of these theories with its overarching concepts. PRS and SCT don't even perform on their original levels when they are made 'relevant'. Since the original tasks were just text in the eyes of RT, adding relevance elements should have made led to at least equal performance levels. RT does not yet replace either PRS nor SCT as a theory on the content effect and this study does not show any real preference for any of the two theories either.

The results of the Drinking Age task show that relevance and/or facilitation is best achieved through simplicity itself. When subjects immediately know the correct selections in any WST, the content effect is at its strongest. Perhaps the quest for the atomic level of facilitation in the content effect has surpassed the practical scope of human reasoning: things like relevance through decreased cognitive effort may work, but they might work best in a small task. The bigger the task, the more facilitators and inhibitors of performance could come into play and the less clear it will be what is actually studied. If anything, future studies on the methods of Relevance Theory should restrict themselves to multiple small tasks with very few varying elements, rather than few larger tasks with multiple elements of relevance.

When designing facilitating effects into WST's, great care must be taken and progress is going to be slow and steady. With many studies using different methods in proving their respective ideas and theories, Relevance Theory can provide a framework to create more comparable tasks. However, the effects of creating a WST with multiple facilitating effects in a Relevance framework could lead to the error of using too many different effects in one method. With the amount of studies and facilitating effects currently available, it would take a long time to learn how exactly which facilitator works on its own and in conjunction with other facilitators. Not to mention the effects of framing the context and the style of narration of a context story, if any.

The herculean task of comparing all possible variations of facilitator combinations and their influences on different groups of subjects worldwide may be too much of a hassle. Yet at the same time, continuously creating new methods to test new ideas and comparing them with totally different methods is not the way to compare results. Perhaps the upcoming and quickly growing field of cognitive neuroscience can provide easier and more efficient models to test with, and lead to more connectionist rather than cognitivist explanations.

Appendix

I. Original Pragmatic Reasoning task, problem 3 from Noveck & O'Brien (1996)

The enriched reasoning-from permission problem:

Below are pictures of four cards containing information about four people. One side of each card shows whether or not a person has taken Action A, and the other side shows whether or not the same person has fulfilled Precondition P.

Suppose that you are an authority checking whether or not these four people are obeying a certain regulation.

The regulation is:

If one is to take Action A, then one must first satisfy Precondition P. In other words, in order to be permitted to do Action A, one must first have fulfilled Prerequisite P.

In order to check whether the regulation has been violated, which of the cards would you turn over? Turn over those cards, and only those cards that you need to check to be sure.

(The four cards show “has taken Action A”, “has not taken Action A”, “has fulfilled Precondition P”, and “has not fulfilled Precondition P”.)

II. Original Social Contract Theory task: Cassava root (Cosmides, 1989)

You are a Kaluame, a member of a Polynesian culture found only on Maku Island in the Pacific. The Kaluame have many strict laws which must be enforced, and the elders have entrusted you with enforcing them. To fail would disgrace you and your family. Among the Kaluame, when a man marries, he gets a tattoo on his face; only married men have tattoos on their faces. A facial tattoo means that a man is married, an unmarked face means that a man is a bachelor.

Cassava root is a powerful aphrodisiac-it makes the man who eats it irresistible to women. Moreover, it is delicious and nutritious-and very unlike cassava root, molo nuts are very common, but they are poor eating-molo nuts taste bad, they are not very nutritious, and they have no other interesting “medicinal” properties.

Although everyone craves cassava root, eating it is a privilege that your people closely ration. You are a very sensual people, even without the aphrodisiacal properties of cassava root, but you have very strict sexual mores. The elders strongly disapprove of sexual relations between unmarried people, and particularly distrust the motives and intentions of bachelors.

Therefore, the elders have made laws governing rationing privileges. The one you have been entrusted to enforce is as follows:

“If a man eats cassava root, then he must have a tattoo on his face.”

Cassava root is so powerful an aphrodisiac, that many men are tempted to cheat on this law whenever the elders are not looking. The cards below have information about four young Kaluame men sitting in a temporary camp; there are no elders around. A tray filled with cassava root and molo nuts has just been left for them. Each card represents one man. One side of a card tells which food a man is eating and the other side of the card tells whether or not the man has a tattoo on his face.

Your job is to catch men whose sexual desires might tempt them to break the law -if any get past you, you and your family will be disgraced. Indicate only those card(s) you definitely need to turn over to see if any of these Kaluame men are breaking the law.

The four cards read: “eats cassava root”, “no tattoo”, “eats molo nuts”, “tattoo”.

III. The Drinking Age task (Griggs & Cox, 1982)

On this task, imagine you are a police officer on duty. It is your job to ensure that people conform to certain rules. The cards in front of you have information about four people sitting at a table. On one side of a card is a person’s age and on the other side of the card is what the person is drinking. Here is a rule: “If a person is drinking beer, then the person must be over 19 years of age”.

Select the card or cards that you definitely need to turn over to determine whether or not the people are violating the rule.

Four note cards labeled, respectively, Drinking a beer, Drinking a coke, 16 years of age, and 22 years of age were presented along with the instructions and rule card.

IV: the relevance and irrelevance conditions of the PRS task (in Dutch)

PRS relevance condition:

Twee professoren aan een universiteit, Professor Jordan en Professor Barkley, zijn bezig met het bestuderen van vier dubbelzijdige kaarten. Op deze kaarten staat informatie over het gedrag van vier personen die aan een onderzoek meegedaan hebben.

Elke kaart laat op de ene kant zien of iemand wel of niet "Actie A" uitgevoerd heeft, terwijl er op de andere kant van de kaart staat of iemand wel of niet aan "Voorwaarde P" voldaan heeft.

Professor Jordan heeft zijn eigen theorie bedacht, en hij denkt dat de data uit het onderzoek zijn theorie absoluut en onvoorwaardelijk zal ondersteunen! Hij stelt dat:

"Als iemand Actie A wil uitvoeren, moet deze persoon eerst aan Voorwaarde P voldaan hebben."

Met andere woorden: je moet eerst aan Voorwaarde P voldaan hebben voordat je Actie A uit mag voeren.

Professor Barkley heeft zo zijn twijfels over de theorie van Professor Jordan. Hij kan zich nog wel een aantal keren herinneren waarbij zijn collega het ook al fout had, wanneer het over zulke zaken ging. Beide professoren zijn te koppig om de vier kaarten zelf te controleren omdat ze allebei geen ongelijk willen hebben.

Jij bent een assistent van Professor Barkley. Je vindt beide professors even aardig en zij hebben altijd al je oprechte mening gewaardeerd. Professor Barkley heeft nu aan jou gevraagd of je de kaarten wil controleren om te zien of Professor Jordan het fout had met zijn theorie.

PRS irrelevance condition:

Twee professoren aan een universiteit, Professor Jordan en Professor Barkley, zijn bezig met het bestuderen van vier dubbelzijdige kaarten. Op deze kaarten staat informatie over het gedrag van vier personen die aan een onderzoek meegedaan hebben.

Elke kaart laat op de ene kant zien of iemand "Actie A" uitgevoerd heeft, terwijl er op de andere kant van de kaart staat of iemand aan "Voorwaarde P" voldaan heeft.

Professor Jordan heeft zijn eigen theorie bedacht, en hij denkt dat de data zijn theorie absoluut en onvoorwaardelijk zal ondersteunen! Hij stelt dat:

"Als iemand Actie A wil uitvoeren, moet deze persoon aan Voorwaarde P voldaan hebben."

Met andere woorden: je moet aan Voorwaarde P voldaan hebben voordat je Actie A uit mag voeren.

Professor Barkley is er niet zo zeker van dat zijn collega gelijk heeft en zijn theorie klopt.

Jij bent een assistent van Professor Barkley. Je vindt beide professors even aardig en zij hebben altijd jouw oprechte mening gewaardeerd. Professor Barkley heeft nu aan jou gevraagd of je de kaarten wil controleren om te zien of Professor Jordan gelijk heeft.

Voor je liggen vier kaarten met de ene of de andere kant naar boven.

V: the relevance and irrelevance conditions of the SCT task (in Dutch)

SCT relevance condition:

Op het afgelegen Polynesische eiland Maku woont de Kulumae stam. Deze stam staat bekend om zijn sensuele cultuur: het is hier vanzelfsprekend om naakt te dansen en rituelen uit te voeren.

Op Maku eiland zijn er echter maar twee soorten voedsel die groeien in het kale landschap: Koriko wortel (een eetbare boomwortel) en Molo noten. De Koriko wortel is moeilijk te vinden op het eiland, maar het wordt gezien als een delicatessen en bovendien gebruikt als een zeer effectieve seksuele lust opwekker! Molo noten daarentegen smaken verschrikkelijk vies, maar zijn wel veel makkelijker te vinden dan de Koriko wortel.

Omdat de Koriko wortel relatief zeldzaam is, wordt het gezien als luxe etensoort die alleen getrouwde Kulumae mannen mogen eten. Hoewel het moeilijk te vinden is, groeit er genoeg Koriko wortel op het eiland om alle getrouwde mannen met gemak te kunnen voeden; getrouwde mannen hoeven dus nooit meer Molo noten te eten als ze getrouwd zijn.

Als een Kulumae man trouwt, krijgt hij uitgebreide tatoeages op zijn gezicht in een speciaal ritueel: zo wordt zijn toetreding tot het getrouwde leven gevierd. Op Maku eiland lopen er echter niet alleen maar mannen met en zonder tatoeages rond. In de afgelopen tien jaar is het bij de Kulumae de mode geworden om een snor te laten groeien. Veel mannen dragen tegenwoordig een snor, en getrouwde mannen met een snor en een tatoeage worden gezien als rolmodellen.

De stamoudsten van het eiland zorgen er voor dat iedereen zich aan eeuwenoude regels houdt. Een van deze regels is dat alleen getrouwde mannen Koriko wortel mogen eten, terwijl ongetrouwde mannen Molo noten moeten eten. Sex voor het huwelijk is eveneens verboden. De stamoudsten letten er vooral op dat ongetrouwde mannen geen Koriko wortel eten en, als gevolg van de lustopwekkende werking, per ongeluk iemand zwanger maken. Het niet naleven van deze regels betekent jarenlange schaamte voor de man en zijn hele familie.

Twee antropologen, Dr. Péron en Dr. Boas, hebben de mannen van de Kulumae stam 18 maanden lang gevolgd en opgenomen met verborgen cameras. De twee doctors hebben op dubbelzijdige kaarten opgeschreven wat de mannen aten en of ze tatoeages hadden of niet.

Dr. Péron, die een nogal ouderwets van aard is, vertelt Dr. Boas: "Ik weet zeker dat alleen getrouwde mannen Koriko wortel zullen eten. Een vrijgezel zou nooit het risico nemen om verbannen te worden en zijn familie eer schade aan te doen"

Dr. Boas, die een nogal avontuurlijk leven geleid heeft in de Caribbische eilanden, is het niet eens met Dr. Péron: "Ach, alle jonge mannen in elke cultuur doen wel eens dingen die ze niet mogen. Er zullen vast wel wat vrijgezelle jongens zijn die toch Koriko wortel zullen proberen te eten."

Dr. Péron antwoordt: "Nou, daar geloof ik niks van. Maar laten er we om wedden!". Hij denkt dit te kunnen controleren door te kijken of de volgende regel klopt:

"Als een mannelijk lid van de stam Koriko wortel eet, heeft hij een tattooëage op zijn gezicht."

Hieronder zie je 4 willekeurige kaarten uit de collectie van kaarten die voor het onderzoek gemaakt zijn. Op een kant staat wat een Kulumae man gegeten heeft en op de andere kant van de kaart staat of deze man een tatoëage heeft of niet. Je kunt alleen niet zien wat er op de andere kant staat!

SCT irrelevance condition:

Op het afgelegen Polynesische eiland Maku woont de Kulumae stam. Deze stam staat bekend om zijn sensuele cultuur: het is hier vanzelfsprekend om naakt te dansen en rituelen uit te voeren.

Op Maku eiland zijn er echter maar twee soorten voedsel die groeien in het kale landschap: Koriko wortel (een eetbare boomwortel) en Molo noten. De Koriko wortel is moeilijk te vinden op het eiland, maar het wordt gezien als een delicatessen en bovendien gebruikt als een zeer effectieve seksuele lust opwekker! Molo noten daarentegen smaken verschrikkelijk vies, maar zijn wel veel makkelijker te vinden dan de Koriko wortel.

Omdat de Koriko wortel relatief zeldzaam is, wordt het gezien als luxe etensoort die alleen getrouwde Kulumae mannen mogen eten. Hoewel het moeilijk te vinden is, groeit er genoeg Koriko wortel op het eiland om alle getrouwde mannen met gemak te kunnen voeden; getrouwde mannen hoeven dus nooit meer Molo noten te eten als ze getrouwd zijn.

Als een Kulumae man trouwt, krijgt hij uitgebreide tatoëages op zijn gezicht in een speciaal ritueel: zo wordt zijn toetreding tot het getrouwde leven gevierd. Op Maku eiland lopen er echter niet alleen maar mannen met en zonder tatoëages rond. In de afgelopen tien jaar is het bij de Kulumae de mode geworden om een snor te laten groeien. Veel mannen dragen tegenwoordig een snor, en getrouwde mannen met een snor en een tatoëage worden gezien als rolmodellen.

De stamoudsten van het eiland zorgen er voor dat iedereen zich aan eeuwenoude regels houdt. Een van deze regels is dat alleen getrouwde mannen Koriko wortel mogen eten, terwijl ongetrouwde mannen Molo noten moeten eten. Sex voor het huwelijk is eveneens verboden. De stamoudsten letten er vooral op dat ongetrouwde mannen geen Koriko wortel eten en, als gevolg van de lustopwekkende werking, per ongeluk iemand zwanger maken. Het niet naleven van deze regels betekent jarenlange schaamte voor de man en zijn hele familie.

Twee antropologen, Dr. Péron en Dr. Boas, hebben de mannen van de Kulumae stam 18 maanden lang gevolgd en opgenomen met verborgen cameras. De twee doctors hebben op dubbelzijdige kaarten opgeschreven wat de mannen aten en of ze tatoëages hadden of niet.

Dr. Péron, die de neiging heeft om te snel conclusies te trekken, vertelt Dr. Boas: "Ik weet zeker dat sommige van de mannen met een snor wel eens Koriko wortel eten. Mannen met een snor zijn gewoon niet betrouwbaar!". Dr. Boas vertelt zijn collega dat hij zich aanstelt.

Dr. Péron reageert: “Je kan zeggen wat je wil, laten we er om wedden!”. Hij denkt dit te kunnen controleren door te kijken of de volgende regel klopt:

"Als een mannelijk lid van de stam Koriko wortel eet, heeft hij een tatoeage op zijn gezicht."

Hieronder zie je 4 willekeurige kaarten uit de collectie van kaarten die voor het onderzoek gemaakt zijn. Op een kant staat wat een Kulumae man gegeten heeft en op de andere kant van de kaart staat of deze man een tatoeage heeft of niet. Je kunt alleen niet zien wat er op de andere kant staat!

VI: The drinking age task (in Dutch)

Als barman/barvrouw van een populaire kroeg in een grote stad moet je je soms aan bepaalde regels houden. Een van de regels waar je op moet letten is:

“Als iemand bier drinkt, dan moet die persoon 16 jaar of ouder zijn”

Er zijn op dit moment 4 mensen aanwezig in de bar. Deze 4 mensen worden gerepresenteerd door 4 kaarten. Op de ene kant van een kaart staat wat ze drinken, op de andere kant staat hoe oud ze zijn. De kaarten geven juiste informatie weer.

References

- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416.
- Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with Wason Selection Task. *Cognition*, 31, 187-276.
- Eliasmith, C. (2005). Cognition with neurons: A large-scale, biologically realistic model of the Wason task. In B. Bara, L. Barasalou & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* (pp. 624-629). Mahwah, NJ: Lawrence Erlbaum Associates.
- Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.
- Evans, J. St. B. T. (1998). Matching bias in conditional reasoning: do we understand it after 25 years? *Thinking and Reasoning*, 4, 45-82.
- Evans, J. St. B. T., & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology*, 4, 391-397.
- Gigerenzer, G., & Hug, K. (1992). Domain specific reasoning, social contracts, and perspective change. *Cognition*, 43, 127-171.
- Goldvarg, Y., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, 25, 565-610.
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407-420.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard Univ. Press.
- Liberman, N., & Klar, Y. (1996). Hypothesis testing in Wason's selection task: social exchange cheating detection or task understanding. *Cognition*, 58, 127-156.
- Noveck, I. A., & O'Brien, D. P. (1996). To what extent do pragmatic schemas affect performance on Wason's selection task? *Quarterly Journal of Experimental Psychology*, 49A, 463-489.
- Parsell, M. (2006). The cognitive cost of extending an evolutionary mind into the environment. *Cognitive Processing*, 7, 3-10.
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 31-95.
- Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.). *New horizons in psychology*. Harmondsworth: Penguin.

Wason, P. C. (1983). Realism and rationality in the selection task. In J. St. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.

Wason, P.C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.

Wason, P.C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23, 63-71.

Wilkins, M. C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology*, 16, No. 102.

Yachanin, S. A., & Tweney, R. D. (1982). The effect of thematic content on cognitive strategies in the four-card selection task. *Bulletin of the Psychonomic Society*, 19, 87-90.