An Evaluation of Gigerenzer's Criticism on the Heuristics and Biases Program's Base-Rate Neglect Studies

WILLEM VAN DIJK

01-07-2016, 30 ECTS
SUPERVISOR: DR. CONRAD HEILMANN
ADVISOR: DR. IR. MARCEL BOUMANS
NUMBER OF WORDS: 35238

Table of Contents

1. Introduction

1.1 General Introduction

Meet Steve:

Steve is very shy and withdrawn, invariably helpful, but with little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail (Tversky and Kahneman 1974, p. 1124).

What is Steve more likely to be: a farmer or a librarian? [1] Many people would say that Steve is more likely to be a librarian. One explanation for this response is that people only evaluate this case by the degree of representativeness or similarity between the description and the two vocations. In doing so, one neglects the so-called base rates[2] that apply to these vocations. That is, the number of (male) farmers is much higher than the number of (male) librarians: according to Shleifer (2012) it is five times as high in the US. When these base rates are taken into account, one is probably less likely to classify Steve as a librarian.

[1] The original thought experiment included a longer list of occupations: e.g. farmer, salesman, airline pilot, librarian, or physician.

[2] I adopt Tversky and Kahneman's (1982B, p. 153) definition of base-rate neglect, who use the term to cover cases in which base rates are entirely ignored and cases in which they are grossly underweighted.

Cases such as Steve's have been used extensively in experiments designed to test the intuitive use of (Bayesian) statistics[3] among humans. Though similar research with more abstract problems already existed in the 1960's, Amos Tversky and Daniel Kahneman popularized the field from the 1970's onwards using real world cases, thereby initiating the 'heuristics and biases' program. Tversky and Kahneman have claimed that intuitive reasoning does not follow Bayesian statistics (Kahneman and Tversky 1972B, p. 450). Instead, people rely on heuristics which can often be quite useful, but can also lead to severe errors or biases. Mimicking the approach of Kahneman's earlier conducted perceptional illusion studies, the heuristics and biases program initially focused on revealing cognitive biases or illusions. The heuristic that will be of concern in this thesis is the representativeness heuristic. This heuristic makes people select those outcomes that are most representative of the evidence provided, such as signified by the case of Steve (Kahneman and Tversky 1973, p. 237).

The name 'heuristics and biases program' is reminiscent of Tversky and Kahneman's 1972 Science article 'Judgment under Uncertainty: Heuristics and Biases'. Gilovich and Griffin (2002) argue that the central idea of the program is that 'judgment under uncertainty often rests on a limited number of simplifying heuristics rather than extensive algorithmic processing', (p.1). A large part of the initial papers, Gilovich and Griffin continue, can be found in the 1982 volume, *Judgment under Uncertainty: Heuristics and Biases* (Kahneman, Slovic, and Tversky, 1982). In addition to the studies of Tversky and Kahneman, other publications that are part of the heuristics and biases program that will be extensively discussed in this thesis are Lyon and Slovic (1976) and Bar-Hillel (1980), who both draw heavily on the earlier publications by Tversky and Kahneman.

The heuristics and biases program has been criticized. Most notably, Gerd Gigerenzer, a psychologist and historian of statistics, is known for his fierce criticism on the heuristics and biases program. Although he praises the use of real world cases, Gigerenzer worried that the Bayes' theorem could not be applied to these. Moreover, he has criticized the representativeness heuristic for being too vague. These and other criticisms will be put under further scrutiny in the next subsection.

Though Gigerenzer has been mostly critical towards Tversky and Kahneman's version of the program, he at times extended criticism towards the whole heuristics and biases program, attacking Tversky and Kahneman and the heuristics and biases program interchangeably. Gigerenzer and Hoffrage (1995, p. 684) see Tversky and Kahneman as the founders and lead proponents of the heuristics and biases program; though Gigerenzer acknowledges others within the program (1991A, p. 85). In this thesis, I will refer to authors and their specific publications whenever possible to avoid confusion. At other times, I will refer to the heuristics and biases program as an entity on its own. The main reason for this is that I discuss Gigerenzer's criticism, who sometimes refers to the heuristics and biases program as a whole.

The goal of this thesis is to study Gigerenzer's criticism on the base-rate neglect studies conducted within the heuristics and biases program; in particular it will focus on those studies conducted by Tversky and Kahneman. For this purpose, I introduce the following research question:

How should Gigerenzer's claims against the heuristics and biases program's base-rate neglect studies be evaluated?

The focus of the thesis will rely predominately on the articles published by Tversky and Kahneman and Gigerenzer in the period from 1971 till 1996. I begin with 1971 as a starting point, as that year features

[3] A more careful definition of 'intuitions' is provided in section 1.4.3.

Tversky and Kahneman's first publication on representativeness. I take 1996 as an end point as Tversky and Kahneman wrote their last joint work on base-rate neglect in that year. Moreover, Gigerenzer shifted his attention from base-rate neglect studies to studies of rationality from 1997 onwards, as Heukelom notes (2005, p. 21). The focus of my thesis is much in line with existing literature on the debate, such as Heukelom (2005)[4], Samuels, Stich and Bishop (2002), Vranas (2000 & 2001), and Jullien and Valllois (2013). Though most emphasis will be placed on publications by Tversky and Kahneman and Gigerenzer in the mentioned period, I will occasionally draw on earlier and later works and contributions by other authors.

One might wonder how an episode in cognitive psychology that ended two decades ago is still relevant today. The simple answer is that Gigerenzer, one of the main players in the debate, still refers to his criticism on the heuristics and biases program's base-neglect studies in his recent work. In a 2015 article, Gigerenzer has argued that the 'base-rate fallacy' is taken by libertarian paternalists Thaler and Sunstein to provide a rationale for intervention in choice situations. Gigerenzer (2015, p. 364) argues against Thaler and Sunstein's position, which maintains that people 'fail to make forecasts that are consistent with Bayes' rule' (Thaler and Sunstein 2003, p. 176). Referring to his earlier criticism on the heuristics and biases program's base-rate neglect studies, he claims that Thaler and Sunstein's position is untenable. In light of this, my thesis can be taken as an indirect contribution to the debate on libertarian paternalism, as it evaluates whether one of Gigerenzer's claims against libertarian paternalism can be substantiated.

A second reason for focusing on the base-rate neglect debate is that the heuristics and biases program and Gigerenzer's own research program[5] are often portrayed as two opposing camps (Gilovich and Griffin 2002, p. 9). Despite this apparent opposition though, both camps have repeatedly relied on each other's contributions. With my study I hope to show to what extent Gigerenzer has contributed to the heuristics and biases program and to what extent he has dissociated himself from it. The results of my study may enhance the understanding of how substantive the current gulf between the camps is.

The results of my evaluation provide a more nuanced perspective on the debate between Gigerenzer and Tversky and Kahneman than one would expect from some of the heated rhetoric that has been associated with the debate. For example, Gigerenzer has argued that the biases described by Tversky and Kahneman simply do not exist, and that their dependence on the representativeness heuristic as an explanation has led to a 'conceptual dead-end' (1991A, pp. 101-102). Kahneman and Tversky (1996) reply that Gigerenzer has misrepresented their position, adding that 'the refutation of a caricature can be no more than a caricature of refutation' (p. 584).

Firstly, I will show that Gigerenzer's criticisms constitute more than a mere refutation of Tversky and Kahneman's work. Especially when it comes to the design of experiments and the applicability of Bayesian statistics in judgmental problems specifically, Gigerenzer has been surprisingly constructive towards Tversky and Kahneman. Moreover, I will show that the framework that Gigerenzer uses to criticize specific features of base rate neglect studies can be used as a tool to find

---

[4] Heukelom (2005, p. 18) puts a few of Gigerenzer's later papers under the label of his core criticism towards Tversky and Kahneman, such as Hertwig and Gigerenzer (1999). As these papers do not focus on base-rate neglect, they are not extensively covered in this thesis.

[5] This research program appears under different names in the literature, with references such as 'simple heuristics', 'evolutionary psychology', 'ecological rationality' and 'fast and frugal heuristics'. For reasons of simplicity I will refer to just Gigerenzer, although it should be emphasized that a large part of his work discussed here appears in co-authored articles or book chapters.

improvements for studies that Gigerenzer himself has not addressed. Despite some reservations, Tversky and Kahneman have been supportive of most of the experimental studies that Gigerenzer designed and conducted as part of his criticism on prior studies.

Secondly, I will show that Gigerenzer's criticism of the representativeness heuristic does not hurt Tversky and Kahneman, as the criticism is mostly misplaced. A corollary of this finding is that the disagreement between Gigerenzer and Tversky and Kahneman does not so much concern the representativeness heuristic itself. Instead, the disagreement between the authors mainly resides in different preferences for a research strategy. Gigerenzer prefers precise models, whereas Tversky and Kahneman think that a demand for preciseness is too strict. I argue that no fundamental gulf exists between the two camps that would prevent both research strategies from coming up with similar theories on the intuitive use of statistics.

I am not the first to comment on the debate between Gigerenzer and Tversky and Kahneman. In the next subsection, I will discuss the most relevant contributions to the debate. My study contributes to this literature by giving a very detailed account of the empirical and methodological contributions of the authors discussed concerning the intuitive use of base rates. Such an enterprise has not been taken up before and is therefore bound to shed new light on earlier findings.

In the next subsection, I will present five claims that Gigerenzer has posed against the heuristics and biases program. In line with Gigerenzer (1996), I will argue that the claims concerning the applicability of Bayes' theorem and the vagueness of the representativeness heuristic are most faithful to his position.

1.2 Gigerenzer's Claims

Floris Heukelom (2005, p. 18) observes a clear thematic and chronological line in the works of Gigerenzer. Gigerenzer started off as a historian of statistics, which led him to criticize Tversky and Kahneman's research. This criticism was a motive for his work on bounded rationality.

In line with Gigerenzer's interests, the following five claims can be distinguished that he has posed to the heuristics and biases program in the realm of base-rate neglect studies:

1. Probability theory is not about single events
2. Statistics does not speak with one voice
3. Proponents of the heuristics and biases program classify judgmental biases as irrational
4. The laws of probability apply only in well-defined circumstances
5. The representativeness heuristic is only a redescription of the phenomenon of base-rate neglect

I do not assume that these claims give a complete overview of the debate between Gigerenzer and Tversky and Kahneman. Rather, these claims represent the issues concerning base-rate neglect studies that have been discussed most forcefully and in most detail by Gigerenzer, Tversky and Kahneman and later commentators on the debate.

In what follows, I will clarify these claims and explain how my thesis picks up on these. Specifically, I will explain why I will not elaborate on the first three claims in the remainder of my thesis. For these claims, other commentators have already argued that they are not central to the debate between Gigerenzer and Tversky and Kahneman. As I have found no strong reasons to doubt the conclusions reached by these commentators, I have decided not to reconsider their work. My comments on Gigerenzer's first three claims are far from comprehensive and I therefore do not wish to make any bold statements concerning those. The considerations that I do voice are foremost speculative additions to claims made by other authors. What I do want to emphasize is that Gigerenzer's fourth and fifth claim do give reasons for a fruitful analysis.

1.2.1 Probability theory is not about single events

In the judgmental problems designed by Tversky and Kahneman, subjects are typically asked to give an estimate for a single event. This design has been criticized by Gigerenzer who at one point has argued that probability theory is restricted to frequencies and does not deal with single events (Gigerenzer 1991A, p. 88). This view is shared by the frequentist view of probability, which is according to Gigerenzer the dominant school of probability. Single events and frequencies are different categories: it is of no use comparing apples and oranges, so Gigerenzer argues.

A weather forecast may illustrate this point. Assume that meteorologists design forecast algorithms on the basis of past weather circumstances. Furthermore assume that heavy clouds in the morning led to precipitation in the afternoon in seven out of every ten days. Now, upon seeing heavy clouds one particular morning, what are the chances that it will rain in the afternoon? From a frequentist point of view one may now argue that the meteorological model only claims that it rains in seven out of ten comparable instances. It thereby does not by itself give a forecast about the weather situation this specific afternoon.

From a frequentist point of view, probability reflects a property of the world. An example to this is the meteorological fact that it rains seven out of ten times under certain circumstances. From a subjectivist perspective, probability not so much reflects a property of the world. Rather, probability is a mental concept, reflecting the beliefs a subject has. Tversky and Kahneman (1974) speak of the 'inherently subjectivist nature of probability' (p. 1130). From a subjectivist perspective, it is fine to give a probability estimate concerning the weather situation for one specific afternoon. The only restrictions one faces in giving such an estimate are those concerning coherence, which will be discussed in section 1.4.2.

Samuels, Stich and Bishop (Samuels, Stich and Bishop, 2002) have argued that Gigerenzer cannot be wedded to a radical frequentist position that denies that probability theory covers single events. As Samuels, Stich and Bishop (2002) have pointed out, Gigerenzer and Hoffrage 1995's 'How to improve Bayesian reasoning without instruction: Frequency formats' makes no sense if this Bayesian reasoning was meaningless in the first place. 'If it ain't broken, you can't fix it', Samuels Stich and Bishop (2002, p. 17) argue and they conclude that Gigerenzer and his sympathizers cannot uphold the claim that the Bayesian norms (as applied to single events) have no relevance.

In support of Samuels, Stich and Bishop's position one could claim that Gigerenzer's (1991A) claim that probability theory does not deal with single events is inconsistent with his second claim concerning competing statistical theories. To cite Gigerenzer (1991A): 'Future research should use competing statistical theories as competing explanatory models, rather than pretending that statistics speaks with one voice' (p. 104). One might now posit the claim that among such competing statistical theories, not only frequentist theories, but also subjective theories are to be included. To argue for that position, one could point out that it is inconsistent to criticize the heuristics and biases program for applying narrow norms, but thereby limiting oneself to a frequentist view of probability at the same time.

One can doubt whether it is fruitful to make an argument along the lines described in last paragraph. The reason for this it that Gigerenzer does not appear to be strongly attached to the claim that probability theory does not deal with single events. As Samuels, Stich and Bishop (2002) have emphasized, Gigerenzer has elicited subjective probability estimates as well frequency estimates in his studies. Moreover, Gigerenzer (1993, p. 290, 1994 and 1996) retreated from his earlier narrow frequentist position, claiming that he does not want to take sides in the philosophical debates on the nature of probability.

Last paragraph's point is in line with Vranas' position in this debate. Vranas (2000) has extensively argued against Gigerenzer that probabilistic norms can apply to single events. In short, probabilistic norms such as the axioms of probability[6] proscribe that subjective Bayesian estimates should follow certain rules of coherence. These probabilistic norms are abstract, in the sense that they cannot be invalidated by references to content of a particular problem. They are therefore content-neutral: they hold independent of the content of a particular problem.

In his reply to Vranas, Gigerenzer (2001, p. 93) argued that 'content-neutral' norms cannot be applied to particular problems in a content-blind way. Reflecting on this reply, Vranas (2001) argues that he and Gigerenzer are addressing different questions. Whereas Vranas (2000) focused on the validity of probabilistic norms, Gigerenzer seems to be primarily interested in statistical norms. In contrast to probabilistic norms, statistical norms do refer to content. The weather forecast example contained a statistical norm, which stated that under certain conditions one can conclude that it rains in seven out of ten instances. The crucial condition that one has to check here is that these conditions are sufficiently specified such that one can assure that the statistical norm indeed applies. Vranas (2001) argues that it is exactly on this matter that Gigerenzer has focused most of his attention, whereas he has left the appropriateness of probabilistic norms untouched. In the end, both Vranas' and Samuels, Stich and Bishop's accounts give reason to argue that Gigerenzer's position is much closer to claim 4 than claim 1.

1.2.2 Statistics does not speak with one voice

Gigerenzer, Hell and Blank (1988), notice that 'psychologists have generally assumed that statistics spoke with one voice' (p. 513), something which they very much oppose. Gigerenzer and Murray (1987, p. 174) claim that there are simply different philosophical views and theories of probability and statistics, just as there are different theories of geometry. A condition that Gigerenzer and Murray (1987, p. 167) put up to show that intuitive statistical reasoning is in err, as Tversky and Kahneman try to demonstrate, is that different statistical approaches should yield the same estimates. Gigerenzer and Murray (1987, p.168-173) show however that estimates for the Taxicab Problem (discussed in section 2.5) are different when for example Neyman and Pearson's statistical theory is applied to it, as Birnbaum (1983) has shown.

Vranas (2000) has specifically replied to the Gigerenzer and Murray's[7] claim that Birnbaum's alternative treatment of the Taxicab Problem does not show that conflicting statistical norms apply to the Taxicab Problem. It rather shows that one should check which assumptions are made in applying statistical norms. Moreover, Vranas argues that even if competing statistical norms would exist, these norms may very well give the same answers for many problems. As Vranas (2000) puts it, one does not expect a statistical theory to come up with 'weird answers to bread-and-butter problems' (p. 190). Vranas emphasizes that in cases of disagreement on what statistical norm to apply, it cannot be ruled out that one of the conflicting norms is the only appropriate one. Thus even if statistics does not speak with one voice, this does not entail that all theories are equally accurate. All in all, Vranas does not think that the problem Gigerenzer has with supposedly conflicting norms is as substantial as he sketches it to be.

---

[6] These are introduced in section 1.4.2

[7] Whenever I refer to Gigerenzer and Murray, I intend to refer to Gigerenzer and Murray (1987)

There is some evidence that Tversky and Kahneman have deliberatively statistical norms other than the ones derived from subjective probability that they used themselves. Although they do not go as far as Lyon and Slovic (1976) who argued that 'the world operates according to Bayes' theorem' (p. 296), Tversky and Kahneman have labeled their Bayesian approach as the normative theory of: prediction (Kahneman and Tversky 1973), subjective probabilities (Tversky and Kahneman 1977B), evidence (Tversky and Kahneman 1977B) and judgment under uncertainty (Tversky and Kahneman 1983). Based on Kahneman's Nobel Prize interview, one can argue that Tversky and Kahneman have wittingly ignored other statistical theories (Kahneman 2002). In the interview, Kahneman stated that in the 1970s the discipline of psychology and decision analysis seemed very promising in providing guidelines for political and governmental decision-making. Heukelom (2005, p. 20) thinks that this is the reason why Tversky and Kahneman put so much emphasis on the normativity of subjective probability theory. By portraying decision analysis informed by subjective probability theory as the sole normative theory, it was more likely to be picked up as the unambiguous tool for governmental decision making, as Heukelom points out. In the end it turned out however, that (high-placed) governmental decision makers were not interested in the comments provided by decision analysis as studied by Tversky and Kahneman. Thus Tversky and Kahneman's focus on the unique normativity of their Bayesian approach failed to serve its purpose in contributing to the construction of a unique theory or tool for decision-making.

One can wonder, as Gigerenzer and Murray (1987, p. 174) do, whether it is not more rewarding to consult the descriptive accuracy of other statistical theories. This does not seem to be a controversial position. To quote Vranas (2001): 'Like many philosophers[8] I do not see the three main kinds of probability concepts - objectivist, subjectivist, and frequentist - as competing' (p. 107). Thus, Vranas concludes that the different concepts of probability and statistical theories can be used side by side without conflict. In that perspective, Gigerenzer's call for statistical diversity is unlikely to cause much dissent.

1.2.3 Proponents of the heuristics and biases program classify judgmental biases as irrational

Gigerenzer does not think that one can measure rationality using normative models that only focus on the structure of a problem (Gigerenzer, Swijtink, Porter, Daston, Beatty and Krüger, 1989). Gigerenzer et al. (1989) argue that 'a rational mind may be more than the kind of intuitive statistician who mechanically applies the same formula (be it expected utility, Bayes' theorem, or some other) to all contents and in all contexts' (p. 232). Gigerenzer and Murray (1987, p. 179) argue that Tversky and Kahneman adhere to this idea of the mechanical intuitive statistician, for whom the use of Bayes' theorem is rational for all contents and contexts.

One may wonder whether Gigerenzer's criticism really hits Tversky and Kahneman's base rate neglect studies. First, recall Vranas' distinction between probabilistic norms and statistical norms introduced in section 1.2.1. This distinction explains that Bayes' theorem holds independent of the content of a particular problem. It does not follow from this neutrality to content Vranas argues, that the theorem can be applied to problems regardless of content and context as Gigerenzer has it. As a result, one can doubt whether Tversky and Kahneman have ever claimed that Bayes' theorem can be blindly applied to all contexts and contents.

The results of experiments on human reasoning have been interpreted differently by adherents of the heuristics and biases program and their critics, as Samuels, Stich and Bishop (2002)

[8] Cited authors suppressed

observe. The former are pessimistic about the extent of humans are capable of rational decision making, whereas the latter are more optimistic. Contrary to this view, Samuels, Stich and Bishop’s (2002) have argued that there is no substantial disagreement between Gigerenzer and the heuristics and biases program on the extent of human rationality. Samuels, Stich and Bishop argue that both adhere to Stein’s so-called ‘standard picture’ of rationality, which entails that a rational person ‘reasons in accordance with principles of reasoning that are based on rules of logic, probability theory and so forth’ (Stein 1996, p.4). Gigerenzer (2008) has replied to Samuels, Stich and Bishop’s article in his book ‘Rationality for Mortals’, arguing that it is not the quantity of rationality that is of concern, but the quality of rationality:

‘We can easily agree how often experiment participants have or have not violated the truth-table logic or some other logical law in an experimental task. But proponents of the heuristics-and-biases program count the first as human irrationality and the second as rationality. I do not.’ (p. 19)

One can doubt whether the position ascribed to researchers in the heuristics and biases program by Gigerenzer (2008) is accurate. In principle, Tversky and Kahneman only test whether people’s intuitions are in line with norms of subjective probability, not different from what Gigerenzer has done in his own works. These norms are part of what Samuels, Stich and Bishop have labeled as the ‘standard picture’ of rationality. This label might be a bit confusing though, as Tversky and Kahneman have never argued that these norms signify the *only* rational course a subject may undertake. That Tversky and Kahneman have a more encompassing view of rationality, is already seen from their famous 1974 Science article ‘Judgment under uncertainty: Heuristics and biases’. In this article, Tversky and Kahneman (1974, p. 1130) define a rational decision maker as someone who incorporates her knowledge on the subject matter, the laws of probability *and* her own judgmental heuristics and biases into her probability estimate. This view of rationality differs from what Gigerenzer has ascribed to Tversky and Kahneman and might even be close to Gigerenzer’s own views.

In his Nobel Prize interview, Kahneman (2002) once again stresses the point that he has not attempted to prove human irrationality with his work. Although he does argue that he and Tversky have shown that people depart from the ideal of rationality as defined in the context of economic theory, Kahneman claims this does not imply that humans are irrational. It does not follow as Kahneman emphasizes, that people are irrational when they do not adhere to theory’s prescriptions. Instead, he believes that humans are reasonable, prudent agents. Later on Kahneman (2011) reconfirmed his unease with the view that his research with Tversky has demonstrated human irrationality. He emphasizes this research was only meant to show that human judgment is not accurately described by the rational-agent model.

The position on rationality that Gigerenzer ascribes to the heuristics and biases program appears not be in line with Tversky and Kahneman (1974) and especially with Kahneman’s late works. The latter furthermore do not disprove Samuels, Stich and Bishop’s conclusion that Gigerenzer and Tversky and Kahneman do not have different views with respect to rationality. Despite Gigerenzer’s claims otherwise, I therefore see no reason to reconsider Samuels, Stich and Bishop’s argument. For this reason, I put the discussion on rationality aside. I furthermore leave the question of how the views on rationality of Gigerenzer and Tversky and Kahneman are exactly related to others.

1.2.4 The laws of probability apply only in well-defined circumstances

A point of concern for Gigerenzer that has returned in every subsection thus far is the applicability of Bayes’ theorem to specific situations. As is apparent from the discussion between Vranas and Gigerenzer, Gigerenzer puts most of his emphasis on assuring the validity of statistical norms for particular problems. Jullien and Vallois (2014, p. 592) claim that Gigerenzer and Tversky and Kahneman address this issue from different angles. Gigerenzer starts off with studying the judgmental problem and constructs a (statistical) norm in light of the specific properties of the problem. Tversky and Kahneman however, take a norm as given and study it by applying it to different contexts. Both Vranas (2001) and Jullien and Vallois (2014) show that Gigerenzer and Tversky and Kahneman do not necessarily disagree on the validity of a norm in a particular context.

In contrast to above authors, Samuels, Stich and Bishop (2002) point out that the issue of applying probability theory to particular problems is among the real disagreements between Gigerenzer and the HandB program. Samuels, Stich and Bishop do not address their concerns in detail. They only state that Gigerenzer is likely to draw upon different branches of probability, which yields a number of allegedly equally reasonable analyses, but with distinct and equally correct answers. Researchers in the heuristics and biases program though, usually assume that only one correct exists.

In section 2 of this thesis I will study in more detail whether the issue of statistical norms divides Gigerenzer and Tversky and Kahneman. I will show that Gigerenzer criticism towards Tversky and Kahneman is in the end much more nuanced than what some of his rhetoric pretends to be. The way in which Gigerenzer engages with base-rate neglect studies foremost shows that he has had contributive aims towards the heuristics and biases program. For a more detailed outlook, the reader is advised to consult section 1.3.

1.2.5 The representativeness heuristic is only a redescription of the phenomenon of base-rate neglect

The other alleged disagreement that Samuels, Stich and Bishop (2002) have pointed out but not elaborated on in detail, concerns the use of heuristics. This classification is in line with Gigerenzer's own observation in his 1996 paper titled 'On Narrow Norms and Vague Heuristics', in which he labels these two issues as 'the main obstacles' towards understanding the cognitive processes that produce both valid and invalid judgments (Gigerenzer 1996, p. 592). The latter is the acclaimed research goal of Kahneman and Tversky (1996, p. 582), which is shared by Gigerenzer.

The main claim of Kahneman and Tversky (1973) is that people predict by representativeness. That is, 'they select or order outcomes by the degree to which the outcomes represent the essential features of the evidence' (Kahneman and Tversky 1973, pp. 237-238). Recall the example of Steve in the introduction, in which one is tempted to say that Steve is more likely to be a librarian, as his personality description is so much librarian like.

Gigerenzer has subjected the representativeness heuristic to some fierce criticism. In particular, Gigerenzer (1996, p. 594) has argued that the representativeness heuristic only gives a restatement of the phenomenon described as base-rate neglect. This argument yields the following three claims, Gigerenzer argues. Firstly, the representativeness heuristic lacks explanatoriness. Secondly, because precise and falsifiable process models are not specified, an account of how the representativeness heuristic relates to such properties as the content and context of a problem is missing. Thirdly, as the just mentioned properties of a problem are disregarded by Bayesian thinking, these are ignored by the representativeness heuristic as well.

I will show most of Gigerenzer's criticism on the representativeness heuristic is misplaced, as he has not accurately described Tversky and Kahneman's position. Firstly, he has failed to distinguish between the types of representativeness judgments. Secondly, Gigerenzer has failed to acknowledge that because representativeness can be verified empirically, its relation to important properties of a problem is also established empirically. I conclude that Gigerenzer's criticism on the representativeness heuristic does not really hit Tversky and Kahneman. At least, representativeness does not make up the 'very real dispute' that Samuels, Stich and Bishop (2002, p. 32) have claimed it to be. In the end, the alleged dispute on the representativeness heuristic appears to come down to a choice for different research strategies. I conclude that it is not certain whether these strategies may or may not converge in the future.

1.3 An Outlook

In this subsection I will set out how this thesis is structured and what approaches are used in tackling the research question. The reader may feel overwhelmed by the amount of detail she will encounter and this subsection will therefore try to offer some guidance in that respect.

The reader should be warned that I use more historical and more analytical approaches side by side. I believe a historical analysis offers more insight in the project that I am advancing in section 2. In section 2, I will go along the specific criticisms that Gigerenzer has cast in relation to the applicability of Bayes' theorem in the experiments conducted within the heuristics and biases program. One must realize that the heuristics and biases program has been developing over the years. These developments show that experimental designs, as well as theory, have improved over time. Some of the improvements proposed by Gigerenzer towards base-rate neglect studies are similar in kind to the improvements applied by researchers in the heuristics and biases program. These earlier versions of base rate neglect studies are often ignored in the literature. Examples include the urn problem in section 1.4.1, the early Taxicab Problem in section 2.4.1 and Hammerton's studies in section 2.4.2. The historical analysis shows that that both Gigerenzer and the heuristics and biases program have been working towards better designed studies, presenting fairer tests of statistical intuitions.

The historical approach also has a role in the treatment of the representativeness heuristic in section 3. Gigerenzer has claimed that the representativeness heuristic stems from Bayes' theorem and he has attached all kind of consequences to this. The history of Tversky and Kahneman's work however shows that their studies on representativeness *precede* their studies on the intuitive use of Bayes' theorem. In a historical sense representativeness at first was independent of base-rate neglect. Historical precedence or independence does necessarily guarantee conceptual independence though, but it does present an interesting clue towards claims concerning the latter. A careful analysis of Tversky and Kahneman's early experiments on representativeness (as introduced in section 1.4.1), that to do not test for base rate use and the later studies that do test for base rate use (as presented in section 2), shows that these studies focus on different aspects of representativeness. I will argue that Gigerenzer has conflated these different uses. I will furthermore show that this causes major problems for his argument against the representativeness heuristic.

The historian of science should be aware. Although I do strive for historical accuracy, I do not pretend to be comprehensive. The historical points in the end serve to contribute to the analytical analysis. In section 2, this analysis is set up in the following manner. Section 2.1 provides the framework that will classify Gigerenzer's critique, which is conveniently handed down by Gigerenzer himself. In what follows, I use Gigerenzer's own framework as well as Tversky and Kahneman's possible replies to evaluate Gigerenzer's criticism. Specifically, I will deal with Tversky and Kahneman's Tom W. study 2.2, the Engineer-Laywer study in section 2.3 and Section 2.4 will discuss the Taxicab Problem and related problems. Although Gigerenzer has mainly focused on what I call the 'later Taxicab Problem', I will show that his framework can be extended to the other problems discussed in section 2.4. Section 2.5 concludes.

Preceding the discussion of Gigerenzer's criticism on Tversky and Kahneman's studies, I will discuss Tversky and Kahneman's methodology towards studies on representativeness and base-rate neglect in section 1.4.2 and 1.4.3. This discussion helps the reader understand how Gigerenzer's position relates to Tversky and Kahneman's. Moreover, it shows how much leeway Gigerenzer has in criticizing the heuristics and biases program's base rate neglect studies.

Section 3 will be organized as follows. In section 3.1 I will introduce the distinction between judgments of and judgments by representativeness as introduced by Tversky and Kahneman in 1982. Subsequently, I will evaluate Gigerenzer's claim that representativeness is merely a restatement of base-rate neglect in section 3.2. Section 3.3-3.5, discuss whether the alleged implications of Gigerenzer's claim hold. In section 3.3, I argue contra Gigerenzer that representativeness can be explanatory. In section 3.4, I discuss Gigerenzer's criticism towards the determinacy of the representativeness heuristic. I will show that Gigerenzer's criticism misses the point in this respect. In section I will evaluate Gigerenzer's claim that the representativeness heuristic has inherited the main attributes of Bayesian reasoning. I will argue that this claim cannot be defended for all but one of the attributes pointed out by Gigerenzer.

Once again, I would like to warn the reader that this thesis discusses many examples of studies on representativeness and/or base-rate neglect. The reader should be aware that these studies are not used in the same way. Some are central to the analysis, such as the Tom W. study, the Engineer-Lawyer study and the Taxicab study. Others only have illustrative or clarificatory purposes, such as the example of Steve already introduced. Again other examples mainly serve the historical approach I use. A lot of examples are used for multiple purposes, as they are for example referred to in both section 2 and section 3. One type of studies that serves multiple purposes are the urn type problems discussed in next subsection. These studies both serve as an introduction to the applicability of Bayes' theorem in section 2, as well as the discussion of representativeness in section 3.

1.4 Introduction to Tversky and Kahneman's work on representativeness

1.4.1 Urn problems

With the increasing popularity of the recently developed subjective Bayesian probability, cognitive psychologists started with experiments on statistical reasoning concerning single events in the 1960's. The questions posed to the subjects were basic variants of the famous urn problems used in text books, with urns and balls substituted for bags and poker chips. Such an experiment is presented below, based on Phillips and Edwards (1966), who presented subjects with (variations of) the following problem:

Consider you are participating in the following experiment. 10 bags are filled with 100 poker chips. 7 of 10 bags have 60% red chips and 40% blue chips, the other 3 bags have 40% red chips and 60% blue chips. From a randomly sampled bag, 20 chips are successively drawn. Of these 20 chips, 12 are red and 8 are blue. What is the probability that the selected bag is the predominantly red one? (pp. 347-348)

To calculate this probability, one can apply Bayes' theorem (Eq.1). We have two hypotheses, namely H_1, that the selected bag is the predominantly red one and H_2, that the selected bag is the predominantly blue one. We now have the following form of Bayes' theorem:

$$P(H_1|D) = \frac{P(H_1) * P(D|H_1)}{P(H_1) * P(D|H_1) + P(H_2) * P(D|H_2)} \text{ [Eq. 1]}$$

$P(H_1|D)$ is the posterior probability of H_1 being true given the data D. $P(D|H_1)$ and $P(D|H_2)$ are the likelihoods of the data, given that $H_1$and H_2 respectively are true. $P(D|H_1) = \binom{20}{12} *.6^{12} *.4^{8} = .1797$ and $P(D|H_2) = \binom{20}{12} *.4^{12} *.6^{8} = .0355$. One now finds that $P(H_1|D) = \frac{.7*.1797}{.7*.1797+.3*.0355} = .9219$. Test subjects have repeatedly been found to give lower estimates of $P(H_1|D)$ (Phillips 1966, p. 35). This phenomenon was named 'conservatism', implying that subjects fail 'to extract from the data as much certainty as is theoretically implied by Bayes' theorem' (Phillips and Edwards 1966, p. 346).

Edwards (1968) concluded that human beings are conservative processors of fallible information. Notice that this does not have to be an irrational strategy: if one is not sure about the trustworthiness of some additional piece of information, it might be safer to stick closer to the base rates. What is furthermore striking about conservatism is that it is opposite to base-rate neglect in the sense that subjects apparently give more emphasis to the base rates than Bayes' theorem can justify. As I will soon come to show though, these conclusions are unfounded. Specifically, one cannot conclude from an estimate that is conservative in light of Bayes' theorem that subjects reasoned conservatively. Tversky and Kahneman's study of the use of the representativeness heuristic in urn problems will clarify this issue.

Tversky and Kahneman first studies of the use of the representativeness heuristic in Bayes' theorem build on the experiments conducted by Edwards and others. Before discussing those, let me first introduce Tversky and Kahneman earliest research on representativeness which is presented in a 1971 study titled 'The Belief in the Law of Small Numbers'. In this paper, Tversky and Kahneman claim that people believe the law of large numbers to apply to small numbers as well. That is, people expect a small sample to be highly representative of the population from which it is drawn, similar to very large samples. As a result, they also expect two samples drawn from the same population to be more similar than sample theory predicts. Remarkably, even expert judges seem to fall in this trap. For example, 84 participants at meetings of the Mathematical Psychology Group and of the American Psychological Association were asked the following question by Tversky and Kahneman (1971):

Suppose you have run an experiment on 20 subjects, and have obtained a significant result which confirms your theory (z = 2.23, p <.05, two-tailed[9]). You now have cause to run an additional group of 10 subjects. What do you think the probability is that the results will be significant, by a one-tailed test, separately for this group? (p. 105)

The median answer was .85, with only 9 respondents giving an answer between .4 and .6[10], the range representing the most reasonable estimates as Tversky and Kahneman (1971, p. 105) claim. Tversky and Kahneman (1971, p. 109) conclude that people appear to be much less conservative in their estimates than Edwards claimed them to be!

---

[9] These are the results to a two-sided hypothesis test. With such a test one can check whether some characteristic of a sample are different from the same characteristic in a population. *Z* denotes the z-score: the higher the absolute value is, the more the sample deviates from the population. *P* stands for p-value: a p-value lower than 0.05 means that for less than one in twenty samples of the population, one expects to find a similar value of the characteristic. The p-value is inversely related with the z-score.

[10] Z-scores decrease with sample size, explaining the lower estimates. For a detailed analysis, check Tversky and Kahneman (1971, p. 105).

In their next paper (Kahneman and Tversky 1972B), Tversky and Kahneman continue their studies on representativeness, presenting amongst other the results of an experiment covering a classical urn problem. This problem is formulated as follows:

Consider two very large decks of cards, denoted A and B. In deck A, 5/6 of the cards are market X and 1/6 are marked O. In deck B, 1/6 of the cards are marked X and 5/6 are marked O. One of the decks has been selected by chance, and 12 cards have been drawn at random from it, of which 8 are marked X and 4 are marked O. What do you think the probability is that 12 cards were drawn from deck A, that is, from the deck in which most of the cards are marked X? (p. 447)

While this version of the experiment has a sample ration of 8:4 and a sample difference of 4, two other problems had the same sample ratio (4:2 and 40:20), while two other problems had the same sample difference (5:1 and 18:4). In another five problems the population proportion was switched from 1/6 and 5/6 to 1/3 and 2/3 respectively. Each problem was presented to a different group of subjects. The following results were obtained: [11]

Table 1
Subjective Posterior Probability in a Symmetrical Binomial Task for Two Pairs of Populations: p=5/6 and p =2/3

p=5/6			p=2/3		
Sample	Subjective Estimate	Bayesian estimate	Sample	Subjective Estimate	Bayesian estimate
4:2	.70	.9615	4:2	.68	.80
5:1	.83	.9984	5:1	.85	.9412
8:4	.70	.9984	8:4	.70	.9412
18:14	.60	.9984	18:14	.58	.9412
40:20	.70	1	40:20	.70	1

These results confirm the earlier findings of conservativeness, as most probabilities are lower than Bayes' theorem proscribes. The more striking result however is that for both population proportions subjects are insensitive to sample difference. At the same time, subjects tend to attach more weight to higher sample ratios. This is exactly opposite to what Bayes' theorem proscribes, as seen from the third and sixth column. Moreover, subjects fail to attach higher estimates to the higher population proportion, also in violation of Bayes' theorem.

These results confirms Kahneman and Tversky's (1972B), conjecture that 'the subjective probability that a sample has been drawn from one rather than another population is a function of the degree to which the sample is representative of each of the populations' (p. 446). Similar to the experiment in the 1971 paper, people expect a small sample to be highly representative of its parent population and therefore assign a high probability to the draw of 5:1. From the perspective of subjective probability theory, these estimations are nevertheless flawed as is most clearly seen from the fact that the subjective estimates do not increase with the Bayesian estimates. Kahneman and Tversky (1972B, p. 450) draw the conclusion that, contrary to Edwards (1968), man is not a Bayesian at all.

The first thing that the reader should pick up from these early contributions by Tversky and Kahneman is that the representativeness heuristic offers an explanation for the conservative estimates

[11] This is a reproduction of Tversky and Kahneman (1972B, p. 448): Bayesian results are not given in the original table.

given by subjects. As is clear from Edwards (1968), psychologists in the 1960's did not agree on how to account for the subjects' estimates in experiments that tested the intuitive use of Bayes' theorem. I will address this issue in more detail in section 3.3, where I will explain that the 'conservative' results to Edwards' bag and pokerchips problem can be put in terms of base-rate neglect.

A second issue concerns the difficulty of the deck and cards problem. Even though Tversky and Kahneman's judgmental problem is already easier to solve than the one designed by Phillips and Edwards, urn problems remain hard to solve, requiring some tedious calculations as we have seen. As Kahneman and Tversky (1972B, p. 447) comment, most subjects restricted their answers to multiples of .10. Moreover, 10% of the respondents[12] even failed to give estimates higher than .5, despite being told that these estimates should necessarily exceed .5. These results show that people's estimates very much deviate from the Bayesian results. Thus, the results of the experiments do show that intuitive statistics are misguided in urn problems. At the same time though, these results are not very much appalling given the difficulty and artificial character of the problem.

In the deck and cards problem, Tversky and Kahneman have focused on the internal validity of experiments on statistical reasoning. That is, they have shown that Edwards' conclusion that man is a (conservative) Bayesian reasoned is not warranted. Preferably, one does not only want to attain internal validity of an experiment, but also external validity. Specifically, one wants to ascertain that the experimental situation reflects everyday encounters that people have with statistical problems. For sure, a hard to solve deck and cards problem does not allow one to draw firm conclusions about people's everyday intuitive statistical reasoning. For this reason, Tversky and Kahneman shifted their focus towards more realistic problems, a move which is appreciated by Gigerenzer and Murray (1987, p. 150).

The shift in content of the experimental problem brings us to the topic of section 2, namely the applicability of Bayes' theorem. As the reader may ascertain, urn problems were designed to fit with probability theory and Bayes' theorem in particular. But as Gigerenzer et al. (1989) have commented, 'the more the urns and balls are filled with content, the more content- dependent reasoning becomes important in addition to the formulae' (p. 228). As a result, the application of Bayes' theorem to less abstract problems may not be that straightforward. Section 2.1 will discuss this issue in more detail.

1.4.2 Tversky and Kahneman's subjective theory of probability and its methodological implications

In this subsection I would like to explain why Tversky and Kahneman should take Gigerenzer's criticism seriously. There are a number of reasons for this. First and foremost, Tversky and Kahneman's view of subjective probability forces them to do so. Second, Tversky and Kahneman have made improvements in the lines of Gigerenzer's criticism to their experimental designs. Third, Tversky and Kahneman have implicitly argued that they are open for new improvements. This subsection shall address the first reason, concerning subjective probability. The second point shall mainly be addressed in section 2.4, while the third point shall be touched upon in section 1.4.3.

As touched upon in section 1.2.1, Tversky and Kahneman embrace a subjectivist theory of probability as a norm. A subjective theory of probability typically only imposes constraints of coherence on probability estimates. Coherence basically entails that a subject's estimations should respect the basic rules of probability as axiomatized by Kolmogorov (Hacking 2001, p. 58-60, 67 and 165). For example, if a subject believes that a coin can fall either tails or heads up, then the probabilities

[12] These subjects were excluded from the analysis

for these two events should add up to one. In more formal terms, $P(Hand\neg H) = P(\neg H) + P(H) = 1$, with $P(H)$ and $P(\neg H)$, the probability of a hypothesis and of its negation respectively.

Tversky and Kahneman argue though that the criterion of internal consistency of probability estimates is insufficient. Instead of putting up an internally consistent set of probabilities, Tversky and Kahneman also require that probability estimates are compatible with the entire web of beliefs an individual upholds. But how then is this compatibility determined? 'Unfortunately', Tversky and Kahneman (1974) comment, 'there can be no simple formal procedure for assessing the compatibility of a set of probability judgments with the judge's total system of beliefs' (p. 1130).

Ideally, a subject herself should be ready to acknowledge that a probability estimate she gave is inconsistent with her other beliefs. Gigerenzer (1996, p. 593) has wondered how this would be possible, as it entails that one must rely on a person's intuitions to prove that people's intuitions are in violation of the normative theory of probability. In principle, this condition requires that after a subject gives an estimate to a problem that does not correspond to the estimate the experimenter had in mind, he should be willing to accept that he was wrong after being explained how the experimenter got to her estimate.

An analogy with studies of visual illusions may illustrate the procedure sketched in the last paragraph. As touched upon in section 1.1, Tversky and Kahneman's studies on statistical intuitions are based on Kahneman's earlier vision studies (Tversky and Kahneman 1974, p. 1124). Moreover, Kahneman (2003) has argued that perception and intuition in general follow the same rules (Kahneman 2003, p. 1450). He adds that he and Tversky were guided by these ideas from their early research on (Kahneman 2003, p. 1452).

Now, consider the Müller-Lyer illusion in Figure 1. When asked which line of the upper set is longest, people tend to point towards the middle one as this the widest of the three arrows. Upon being confronted with the lower set though, they come to realize that their first estimate was wrong as they will acknowledge that all lines have the same length.

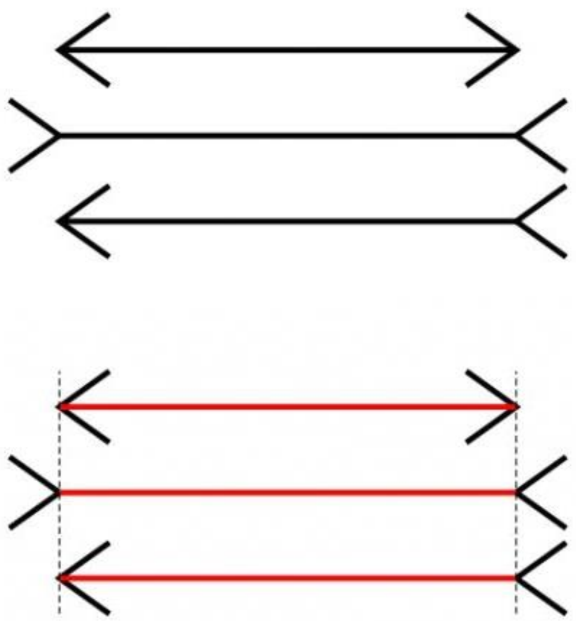

Figure 1

The illusory nature of the Müller-Lyer illusion is often much easier to get across than that of cognitive illusions. For example, in the Tom W. experiment, subjects are asked predict Tom W. graduation area based on a personality description[13], which very much resembles a student of the exact sciences. After the subjects gave their estimates, they were told that Tom W. is enrolled in the Education Program. Subsequently, the subjects are asked to explain the relation between Tom W.'s career choice and his personality. Despite the fact that the subject knew that personality descriptions tend to be unreliable, most of the subject kept to their prior judgments in their explanations. Apparently, subjects find it difficult to revise prior intuitions after being given new information. Instead, they are biased towards rationalizing or confirming their prior intuitions (Tversky and Kahneman 1977B, pp. 2–14). For this reason it is often not possible to judge a person's intuitions by invoking her own intuitions about what the correct approach should be, as in the Müller-Lyer illusion.

One could possible circumvent the confirmation bias described in last paragraph by asking the subjects how their intuitive estimate came about. If the reported strategy deviates from the correct procedure, one may able to convince the subject that the intuitions were wrong. Gigerenzer, Hell and Blank (1988, p. 518) do report such strategies for example.

Kahneman and Frederick (2002, p. 58) criticize this method though. They point out that corresponding to evolutionary history, intuitive judgments occupy a position in between perception and reasoning, with the boundaries between perception and judgment being fuzzy and permeable (Kaheman and Frederick 2002, p. 50). As a result, people only have a weak grasp of how their intuitive judgments came about. This weak grasp is further troubled by the presence of the representativeness heuristic. As Heukelom (2009) has commented, heuristics for Tversky and Kahneman 'do not yield the decision, but reorganize the informational input in such a way that a decision making process is possible' (p. 83). Or as Kahneman and Frederick (2002) argue: 'When confronted with a difficult question people often answer an easier one instead, usually without being aware of the substitution' (p. 53). Thus, the representativeness heuristic makes people focus on different information, without them being conscious of it.

Because of the confirmation bias, it appears that it is often not possible to have the same person being both the subject and the judge of an experiment. In other words, one may not be able to ask the subject whether her estimate is inconsistent with her 'entire web of beliefs' as with the Müller-Lyer illusion and as Tversky and Kahneman's theory of subjective probability ideally has it. As a consequence, one cannot test whether a judgmental problem has an entirely tight single correct estimate to it. For this reason, one might always find some reasons to criticize the design of a problem. Whether such criticism is valid is mostly an empirical question, which might be hard to address, as I have explained above. Nevertheless, one can of course try do evade possible criticism by making the design of a problem as strong as possible. I believe this to be the objective of both Tversky and Kahneman and Gigerenzer.

One can conclude from this subsection that Tversky and Kahneman should take Gigerenzer's criticism seriously in those cases that he can make a credible claim that a subject could have a sensible view on a judgmental problem that differs from that of the experimenter. If such criticism could lead to an improved design of a problem without incurring extra costs, there should be no reason for Tversky and Kahneman not to support such proposals. That a changed design does have its costs will be discussed in coming subsection. A second point that the reader should take from this paragraph is that there not so many clear-cut cases in this debate that one can really press upon the other party.

[13] Consult section 2.3.1 for Tom W.'s personality description

Except for the obvious mistakes, points of critique are debatable as one can never precisely find out how subjects have understood a problem.

1.4.3 Tversky and Kahneman's Methodological Considerations

In their 1982 book chapter 'On the Study of Statistical Intuitions', Tversky and Kahneman reflect on the methodology of their prior research. I will introduce two major issues here. First, I will address the issue of 'Socratic hints', which Tversky and Kahneman bring to claim that information given in a problem should not be too obvious. Second, I will discuss Tversky and Kahneman's concerns about the context of the question-answering paradigm, which may lead subjects to make different inferences than in other contexts.

The goal of Tversky and Kahneman's paper is to argue how an experiment can be a fair test of intuitive judgment. Tversky and Kahneman argue that the terms 'intuition' and 'intuitive' are used in three different ways. First, a judgment is termed intuitive if it is attained through an informal and unstructured way of reasoning, which does not include the use of analytic methods or deliberate calculation. Second, a rule or fact of nature is called intuitive when it corresponds with our lay model of the world. Third, a rule or procedure is part of our intuitions when we apply it in normal conduct.

According to Kahneman and Tversky (1982A) a test of fair intuitions should not provide too many Socratic hints or instructions. With Socratic instruction Tversky and Kahneman mean that problems are formulated in such a way, that subjects are overtly guided to the right answer. This type of instruction resembles Socratic methods of questioning people, as Socrates claimed that people had the proper knowledge all along, but only need to remember it. But although the steps to the right answer might be highly intuitive, the eventual outcome is often not immediately compelling. If a result is not immediately compelling though, it is no longer intuitive, so Tversky and Kahneman argue. For this reason, Socratic instructions should be avoided.

As an example to Socratic instructions, Tversky and Kahneman (1982A) come up with the following case:
'Which hospital -a large or a small one- will more often record days on which over 60% of the babies born were boys?' (p. 130)

Apparently, this has been a difficult question for Stanford undergraduates as Kahneman and Tversky (1972B, p.441) have shown. Tversky and Kahneman (1982A) now argue that through a sequence of small and easy steps, the right answer can be elicited with more success:

'Would you not agree that the babies born in a particular hospital on a particular day can be viewed as a sample?'
'Quite right. And now, would you have the same confidence in the results of a large sample, or of a small one?
'Indeed. And would you not agree that your confidence is greater in a sample that is less likely to be in error?
'Of course you had always known that. Would you now tell me what is the proportion of boys in a collection of babies which you consider the closest to an ideal of truth?
'We agree again. Does that not mean, then, that a day on which more than 60% of babies born are boys is a grave departure from that ideal?'
'And so, if you have great confidence in a sample, should you not expect that sample to reveal truth rather than error?'. Etc. (pp. 130-131)

Kahneman and Tversky (1982A) claim that 'there are no rules that distinguish fair tests of intuitions, from contrived riddles on the one hand, and from Socratic instruction on the other' (p. 130). They do seem to point at some spectrum of questioning, with contrived riddles at the near and Socratic instruction at the far end of it. Fair tests of intuitions are located in the middle, but there are no clearly aligned borders between the different sorts of questioning.

The discussion of Socratic hints show that a change to the design of a problem can have its costs. When the properties of a problem are made more salient, subjects are steered into a certain direction. This is a cost for the experimenter in case she wishes to mimic daily choice situations that a subject encounters, which may often lack manipulatory properties. At the same time, including versions of a problem with manipulations may be of benefit for the experimenter, as this allows her to study how a subject behaves under different circumstances. For this reason, Tversky and Kahneman have been quite sympathetic to Gigerenzer, Hell and Blank's manipulation to random sampling in the Engineer-Lawyer (Engineer-Lawyer) study, as discussed in section 2.4.3.

After discussing Socratic instructions, Tversky and Kahneman address the question-answering paradigm more broadly. With the question-answering paradigm Tversky and Kahneman mean that in a typical experiment on judgments under uncertainty subjects are either verbally or orally presented with pieces of information, about which they have to answer questions or estimate values. The downside of this paradigm is that it is not guaranteed that it mimics the inferences people make under normal circumstances. For example, judgments in everyday conduct are often not made in response to outright questions.

Besides the problem of creating artificial situations, a further problem with the question-answering paradigm concerns the position of the experimenter vis-à-vis the subjects. As subjects know that the problems posed to them are constructed by experimenters, these problems may elicit more thoughts in the subjects than they would have otherwise have. Their consideration of the problem is often not restricted to its content. A subject may for example doubt whether there is a correct answer at all to the question posed. She might furthermore wonder about the selection process of the information or the possibility of inclusion of misleading information. These thoughts are most often not observed, but they do influence the final answer that the subjects give.

A further problem with experimental situations is that they may violate conversational rules, such as the cooperativeness principle. This principle entails that the listener is safe to assume that the speaker is cooperative towards her, thereby trying to be 'informative, truthful, relevant and clear'. The advantage of the cooperativeness principle is that it reduces ambiguities. Thus, when someone says that 'Susan tried to clean the house', it follows that this attempt was not successful.

Tversky and Kahneman conclude that the conversational aspect has not been given sufficient attention in studies of judgment, including their own. Differences in presentation do matter for the interpretation of facts. The rules that guide communication among people simply do not apply when information is gathered through interaction with nature. Finally, it is difficult to avoid giving useful or misleading clues about the correct answer when posing questions, or avoid disclosure of information regarding the expected response.

It appears that Tversky and Kahneman have been very much aware of the problems concerning the external validity of experimental studies of judgment under uncertainty. They emphasize that these issues require more careful attention. What is more, they even acknowledge that their own earlier research has insufficiently done so. I conclude from this that Tversky and Kahneman are in principle very much open to suggestions concerning weaknesses and possible improvements in experimental design. Unsurprisingly, Gigerenzer has provided many of such suggestions, which will be the topic of next section.

2. On the Application of Statistical Norms

2.1 The Isomorphism Conditions

2.1.1 Introducing the Conditions

Recall the example of Steve as introduced in section 1.1. I have argued that one can easily be misled by the high similarity between the description of Steve and the stereotype of the librarian, thereby overlooking the low frequency of librarians in the population. Despite of this however, it is not entirely clear whether one commits a fallacy in general if one thinks that Steve is most likely to be a librarian.

Let me address a few issues about Steve's case to illustrate the point that Bayes' theorem cannot be applied univocally to a specific problem. Firstly, one might simply not depend on the USA base rates, which tell that the number of farmers is five times as high as the number of librarians. Instead, one may rely on one's own peer group, which may contain relatively more librarians. Secondly, one may not believe that 'Steve' is randomly sampled from the population, but that the description is deliberately constructed to resemble a librarian. It would be misleading or contradicting to conversational rules to give a description that is very similar to a librarian, while not aiming to refer to a librarian (cf. Kahneman and Tversky 1982A, p. 133). As a result, you conclude that base rates should be neglected.

Thirdly, in the same vein, you might have been so much caught by the similarity of the description to a librarian that you come to misunderstand the question to ask whether Steve is more like a librarian than whether he is more likely to be a librarian. Fourth, you might reason that the two events, being a farmer or a librarian, are not mutually exclusive. As this is a requirement for the use of Bayes' theorem, you conclude that you should use another way to determine which occupation Steve is more likely to practice. Given either more emphasis to the resemblance of Steve's description or to the base rates, you conclude that Steve is more likely to be a librarian or a farmer respectively. Fifth, you may simply happen to think very drastically about the resemblance of the description, such that you indeed revise your prior beliefs that only relied on the base rates to the contrary.

Most of Gigerenzer's criticism towards the design of Tversky and Kahneman's experiments is cast along the same lines as the remarks on the case of Steve. Gigerenzer and Murray (1987, p. 162-167) have elaborately discussed the difficulties that one may run into when trying to construct or answer problems using Bayes' theorem. According to Gigerenzer and Murray an 'isomorphism assumption' makes sure that problems can be unambiguously mapped into Bayes' theorem. In other words, the isomorphism assumption makes sure that real world problems sufficiently mimic the unambiguity of text-book problems such as the urn problems discussed in last subsection. Gigerenzer and Murray distinguish two types of isomorphisms, which are both regarded as necessary conditions for the isomorphism assumption.

The first type is the concept isomorphism, which ensures that 'each of the two formal concepts in Bayes' theorem is unequivocally matched with one semantic concept concerning the problem' (Gigerenzer and Murray 1987, p. 163). These two formal concepts concern base rates and (conditional) likelihoods in Bayes' theorem. It should be clear for the subject which base rates and conditional likelihoods apply to the problem at hand. In case these are explicitly given, the subject should have sufficient reason to judge these as relevant to the problem. Concept isomorphism ensures that there is only one candidate for each concept.

The second type is structural isomorphism, which ensures that 'the formal structure underlying Bayes' theorem is represented by a similar structure of the problem' (Gigerenzer and Murray 1987, p. 163) . First, the evidence E should be randomly sampled from the population to which the base rates refer. Second, the use of Bayes' theorem necessitates that the hypotheses are mutually exclusive and exhaustive[14]. Third, consecutive samples taken from the same population should be mutually independent. Thus, in an urn problem, when multiple draws are made from the same urn, balls should be put back after each draw to secure independence.

Although Gigerenzer in his later work only sparsely explicitly refers to the isomorphism conditions, his criticism can be easily mapped into this framework. For this reason I will use these conditions as a point of departure for my evaluation of Gigerenzer's claims concerning Tversky and Kahneman's studies. I will use Gigerenzer's criticism to construct the views that a subject may have on the judgmental problems presented by Tversky and Kahneman. Specifically, a subject may have reasons to think that Bayes' theorem does not apply. Consequently, I will discuss how Tversky and Kahneman would go about these lines of thought. Crucially, Tversky and Kahneman are to take these seriously, given their commitment to a subjective view on probability.

Table 2

Conditions for the Isomorphism Assumption

Concept Isomorphism	**Structural Isomorphism**
1) Base rates	4) Random sampling
2) Conditional Likelihoods	5) Mutual Exclusiveness
3) Posterior probability	6) Independent Drawing

Although Gigerenzer has left it unmentioned, I would like to include another condition among the concept isomorphisms, namely posterior probability. In a sense the posterior probability is also a formal concept in Bayes' theorem that needs to be matched to a semantic concept. The point is that this semantic concept is not necessarily unambiguous from the perspective of the subject. For example, a subject may not understand a question that is supposed to elicit a probability estimate as such because the wording of the problem may lead her to think otherwise. Gigerenzer (1996, p. 593) has cast this type of criticism to the famous Linda problem and we will see that a similar point of criticism can be posed towards the Tom W. study. Although his treatment of the Linda problem provides reason that he would agree, it is not so relevant whether Gigerenzer would actually agree on adding the extra condition. I only invoke this condition to summarize the criticism towards Tversky and Kahneman's studies that deals with the posterior probability. Moreover, as I will show in section 2.4.2., the posterior probability condition allows one to point out weakness in experimental designs that have so far not been mentioned in the literature.

I believe that Gigerenzer's aims with the formulation and elaboration of the isomorphism conditions are at least twofold. First, his aim is to show that the conditions are restricting the applicability of Bayes' theorem to (real world) problems, opening the way for other approaches to tackle these. Secondly, his aim has been constructive, attempting to improve or add otherwise to the design of Tversky and Kahneman's studies, as to better capture people's statistical intuitions. This second aim has been most pronounced in the articles covered in this thesis. This should come as no

[14] Simply put: $P(H\&\neg H) = P(\neg H) + P(H) = 1$, with H being some hypothesis

surprise, as Gigerenzer has conducted series of studies involving the intuitive use of Bayes' theorem, thereby accepting the validity of the Bayesian framework for the judgmental problems presented.

2.1.2 A Preview

In the coming sections I will discuss Gigerenzer's criticism towards Tversky and Kahneman's base-rate studies. More precisely, I will criticize the studies from the adapted Gigerenzerian isomorphism framework, including points of criticism that Gigerenzer has made in other contexts, but which can applied to the studies discussed in this thesis as well.

For a quick overview, one can consult Table 3, which shows whether the isomorphism conditions are satisfied in the different studies discussed. Notice that because Gigerenzer has not criticized the Tversky and Kahneman's studies covered in this thesis from the perspective of the isomorphism conditions of independent drawing and mutual exclusiveness, I will ignore these from here. Table 3 shows from which isomorphism conditions Gigerenzer has criticized the studies. A 'V' denotes that Gigerenzer has not criticized the study from a certain isomorphism condition. A '?' signifies that Gigerenzer has criticized the study, but that this criticism has not been (fully) acknowledged by Tversky and Kahneman. An 'X' shows that the heuristics and biases program at some point acknowledged that a condition was not fulfilled, regardless whether Gigerenzer was the first to criticize it.

In the next three sections, I will discuss three of Tversky and Kahneman's most famous base-rate neglect studies. The set-up for these sections will be as follows. First I will introduce the study as Tversky and Kahneman have initially presented it. Secondly, I will discuss the points of criticism that Gigerenzer has given or could give from his isomorphism framework towards these studies. Thirdly, I discuss Tversky and Kahneman's (possible) response to Gigerenzer's criticism. Section 5 on the Taxicab Problem deviates a bit from this approach. In section 5, I dwell upon a number of other studies on base-rate neglect. In this section, I invoke the Suicide and the Intercom problem to evaluate Gigerenzer's criticism towards the Taxicab Problem. Moreover, I refer to the Hammerton's Medical Screening Test and the Engine Crack Test, to show that the heuristics and biases program has been involved in improving experimental design.

Table 3

Conditions for the Isomorphism assumptions

	Base rates	**Conditional Likelihoods**	**Posterior probability**	**Random sampling**
Deck and Cards	V	V	V	V
Tom W (1)	V	V	?	X
Tom W (2)	V	V	?	?
Engineer-Lawyer (Tversky and Kahneman)	V	V	V	?
Engineer-Lawyer (Gigerenzer, Hell and Blank)	V	V	V	V
Cab (Early)	?	X	V	?
Cab (Later)	?	V	V	?

Suicide	V	V	V	V
Medical Screening	V	V	?	?
Engine Crack	V	X	?	V
Engine Crack II	V	V	?	V
V = Not criticized by Gigerenzer *? = Criciticized by Gigerenzer: criticism not (fully) accepted by heuristics and biases program* *X = Conditions of which the heuristics and biases program has admitted they are not fulfilled*				

2.2 Tom W.

2.2.1 Introducing Tom W.

The Tom W. study, which was first published in 'On the Psychology of Prediction', Tversky and Kahneman have tested how subjects deal with representative though fallible information (Kahneman and Tversky 1973, p. 238-241). In this between-subject study, a first group of 69 subjects was asked to give the percentage of first year graduate students in the U.S. enrolled in the nine fields of specialization as listed in Table 4. The mean estimates of these base rates given by this 'base-rate group' are given in the second column.

Table 4
Estimated Base Rates of the Nine Areas of Graduate Specialization and Summary of Similarity and Prediction Data for Tom W.

Graduate Specialization Area	Mean judged base rate (in %)	Mean similarity rank (1-9)	Mean likelihood rank (1-9)
Business Administration	15	3.9	4.3
Computer Science	7	2.1	2.5
Engineering	9	2.9	2.6
Humanities and Education	20	7.2	7.6
Law	9	5.9	5.2
Library Science	3	4.2	4.7
Medicine	8	5.9	5.8
Physical and Life Sciences	12	4.5	4.3
Social Science and Social Work	17	8.2	8.0

Another group of 65 subjects, the 'similarity group', was given the following personality sketch by Kahneman and Tversky (1973):

Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense. (p. 238)

This group was asked to rank the nine specialization areas in terms of the similarity of Tom W. to the typical graduate students in these fields from 1 (most similar) to 9 (least similar). The mean similarity ranks are given in Table 4. Finally, Kahneman and Tversky (1973) a last group, the 'prediction group', consisting of 114 US graduate students in psychology the personality sketch and the following additional information:

The preceding personality sketch of Tom W. was written during Tom's senior year in high school by a psychologist, on the basis of projective tests. Tom W. is currently a graduate student. Please rank the following nine fields of graduate specialization in order of the likelihood that Tom W. is now a graduate student in each of these fields. (p. 239)

The results are given in the fourth column of Table 4. Surprisingly, mean similarity ranks and mean likelihood ranks are highly correlated, with a correlation coefficient of .97, whereas the correlation coefficient for estimated base rates and likelihood rank is -.65[15]. One would have expected the subjects to be more conservative given the unreliability of personality sketches. When asked about the accuracy of projective tests in predicting the first choice among the nine areas, the third group gave an estimate of 23% of correct predictions. Despite of this awareness, the subjects failed to be conservative with respect to the base rates in the likelihood ranking.

Apparently, the prediction group turned out to have performed the same task as the similarity group, although they were asked a different question. As an explanation, Kahneman and Tversky (1973) argue that 'people predict by representativeness, that is, they select or order outcomes by the degree to which the outcomes represent the essential features of the evidence' (pp. 237-238). Kahneman and Tversky (1973) conclude that the accompanying disregard for prior probabilities is 'perhaps one of the most significant departures of intuition from the normative theory of prediction' (p. 243).

2.2.2 Concerning the isomorphism conditions

The special feature of the Tom W. study is that explicit information on base rates and conditional likelihoods are not given, such that no unique solution can be given to this problem. This of course does not mean that Bayesian reasoning cannot be used to assess this problem, as it still makes sense to compare base rates with similarity ranks and the predictions. Gigerenzer has not criticized the Tom W. study in light of the base rate or the conditional likelihood assumption. Apparently he has accepted the assumption that the similarity rank and the base rates which were elicited in the first two subject groups, were also implicitly embraced by the prediction group. Gigerenzer directed his criticism towards the condition of random sampling. Moreover, I will argue that his criticism on the polysemy of the famous Linda problem can be extended to the Tom W. study. I will discuss this issue first.

As already touched upon, the second and third column of Table 4 do give a clear indication of the subject's knowledge of base rates and conditional likelihoods. At the same time though, it is not so clear whether the concept of posterior probability is translated into an unambiguous concept. Specifically, one may doubt whether the subjects correctly understood whether they were asked about the posterior probability or the similarity rank instead.

Although Gigerenzer has not discussed the ambiguity of the wording in the Tom W. problem, he has done so for the famous Linda problem together with Ralph Hertwig (Hertwig and Gigerenzer 1999, p. 277). Specifically, they argue that the use of the term 'probable' is not restricted to the realms of mathematical probability, but also refers to plausibility, as the Oxford English Dictionary tells. Furthermore, when one studies the etymology of similarity and probability Hertwig and Gigerenzer

[15] Likelihoods were inverted in the calculation of this coefficient

argue, one finds that both terms initially had the same meaning. They even speculate that this common root drives the representativeness heuristic.

The OED defines likelihood as 'the quality or fact of being likely or probable' (Likelihood, 2016). Thus, likelihood could be understood as plausibility depending on the context of the problem. Hertwig and Gigerenzer stress that it is the subject who has to decide, given the specific situation, which meaning is referred to. Upon asked what is the most plausible scenario, subjects will evaluate this question by similarity as such a judgment would make a more credible story. Buturovic and Tasic (2015, p. 133-134) have also pointed out that Gigerenzer's criticism can easily be extended to Tom W. They argue that the question can be interpreted in the following way: 'Who is Tom more like—a computer scientist or a scholar of English poetry?'

It appears that the wording of the Tom W. problem does not single out whether posterior probability or conditional likelihood is asked for. In case of the Linda problem, Gigerenzer has argued that it is a sign of intelligence that people know how to flesh out the meaning of polysemous words like 'probable'. To call this a cognitive illusion would be misplaced. Notice that subjects will have more reason to think that the latter is asked for when the condition of random sampling is not satisfied, an issue that I will now further address.

Gigerenzer has repeatedly criticized the Tom W. study for not satisfying the condition of random sampling. (Gigerenzer and Murray 1987, p. 166, Gigerenzer et al. 1989, p. 231, Gigerenzer 1991A, p. 96, 1991B, p. 262, 1993, p. 309). As the Tom W. problem does not convey information about random sampling of the personality sketch, it is justifiable that the subjects neglected the base rates. Without random sampling, base rates do not apply and the only information one can rely on is that concerning similarity. This explains why the similarity and prediction ranks were so similar.

Another reason why subjects were led by similarity instead of base rates is that the personality sketch was constructed to bear higher resemblance to a typical student in one of the smaller fields (computer or library science), as Kahneman and Frederick (2002, p. 62) have acknowledged. When subjects have reason to doubt that the personality description is not randomly picked, it may be more worthwhile to doubt whether base rates apply. Gigerenzer, Hell and Blank (1988, p. 519) refer to a German TV program in which experts have to guess the profession of candidate, who they can ask yes-no questions. Because these candidates were selected and not randomly drawn, the experts know that they better not revise their opinions through Bayes' theorem, but anticipate encountering rare professions from the start. Although Gigerenzer only made this argument in the context of the Engineer-Lawyer study, I believe it also applies to the Tom W. study, as Kahneman and Frederick have acknowledged that the personality description was purposely artificial.

2.2.3 Tversky and Kahneman's reply

I have shown that the condition on posterior probability and random sampling can be criticized from a Gigerenzerian point of view in the Tom W. studies. In this subsection I will discuss how Tversky and Kahneman would react to these points of criticism. I will start off with random sampling and deal with posterior probability afterwards.

Although Gigerenzer's criticism does make sense for the first Tom W. study, it loses force against the second study in which random sampling is explicitly mentioned, which was emphasized in Tversky and Kahneman's reply to Gigerenzer's critique (Kahneman and Tversky 1996, p. 585). For some reason, Gigerenzer has repeatedly failed to acknowledge this, even in a post-1996 article (Gigerenzer 2015, p.372). In the second study, Tversky and Kahneman present five personality sketches of ninthgrade boys[16] to the prediction group, supposedly randomly sampled from a longitudinal study. The results for this study were very similar to the first, with high correlations between similarity ranks and likelihood ranks and low correlation between likelihood ranks and base rates. Apparently, it did not matter whether it was mentioned that the description was randomly sampled.

Gigerenzer would argue against the second Tom W. study that the condition for random sampling is still not secured. His second argument against still stands. Even if it were mentioned that the description was randomly sampled, subjects would be right if they would think that they are fooled. And as Kahneman and Tversky (1982A, p. 132) claim, subjects are usually concerned about the possibility of being misled by the information. So if pressed on this, Tversky and Kahneman should agree that the condition of random sampling was not satisfied for the subjects thinking along these lines.

Tversky and Kahneman might counter Gigerenzer's criticism by arguing that it would be against the cooperative principle to say that Tom W.'s personality description was randomly sampled when in fact it is not. This gives reason to argue that if subjects were to distrust the information on random sampling, this probably has been a small group. At least, it would be hard to defend that the fact that the similarity and prediction group gave nearly equal answers is entirely driven by the subjects' distrust in the random sampling of the personality description. In the next section on the Enginer-Lawyer study, I return to this issue of random sampling.

To classify the representativeness heuristic as a bias one needs a clear distinction between similarity and probability. Ideally one would like to know how the subjects came to give probability estimates (allegedly) only based on similarity and not on base rates. Was it just the wording of the problem that led the subjects to answer a different question than the studies thought they had posed? Or was it the high similarity between the personality description and a typical student of some of the Graduate Specialization Areas that made the subject substitute the question about prediction for one about similarity? Or was it a combination of both perhaps?

As long as the meaning of words as probability is not as clear cut in natural language as in mathematical language, one may find reasons to argue about the biasedness of the representative-ness heuristic. As the questions I posed in last paragraph are difficult if not impossible to answer, such discussions are bound to be indecisive.

What have been more fruitful from an empirical perspective are the studies that have been able to circumvent the issue of polysemy. For example, studies using frequency formats do not use polysemous terms. McCauley and Stitt (1978, p. 929) have conducted a study on stereotypes in reply to the Tom W. study. Instead of predicting from personality to category, McCauley and Stitt took a reversed approach. Participants were asked to give 'the percent of Germans who are efficient' $P(H_1|D)$. McCauley and Stitt invoke Bayes' theorem, which gives $P(H_1|D) = \frac{P(H_1)*P(D|H_1)}{P(D)}$. They furthermore ask the subjects to give "the percent of all the world's people who are efficient" $P(H_1)$, 'the percent of efficient people who are German' $P(D|H_1)$ and 'the percent of the world's people who are German' $P(D)$. With these last three estimates, one can calculate the accompanying posterior probability $P(H_1|D)$. A comparison with the directly estimated $P(H_1|D)$, showed that subjects performed much in line with Bayes' theorem, not equating $P(D|H_1)$ with $P(H_1|D)$ as the representativeness heuristic describes. Similar results with frequency formats have been found by Gigerenzer and Hoffrage (1995).

[16] For clarity, these concern 14 years old boys.

Tversky and Kahneman (1996, p. 587) have not disqualified the use of frequency formats instead of probability formats. They have only emphasized that such a change is analogous to adding the dashed vertical line in the Müller-Lyer illusion (Figure 1). Using frequency formats simply makes a judgmental problem less opaque, a conclusion that Gigerenzer and Hoffrage (1995) are probably willing to agree on, arguing that 'the mind is tuned to frequency formats' (p. 697).

Other ways out of the potential problems with polysemy are provided in the next sections. The Engineer-Lawyer study, which in Tversky and Kahneman's words provide a 'more stringent test' for the representativeness heuristic and the neglect of base rates, does not face the same problems with polysemy as the base rates are made 'exceptionally salient' and compatible with the response mode (Kahneman and Tversky 1973, p. 241). Another way out of the problem of polysemy is through judgmental problems that do test for the use of base rates, but not focus on representativeness specifically. In those contexts, a possible ambiguity of words such as 'probability' is less likely to arise. These studies will be discussed in section 2.4, while next section will focus on the Engineer-Lawyer study.

2.3 The Engineer-Lawyer Study

2.3.1 oducing the Engineer-Lawyer study

The Engineer-Lawyer study was first presented by Tversky and Kahneman in their 1973 article, 'On the Psychology of Prediction'. The major difference between Engineer-Lawyer and Tom W. is that prior probabilities are explicitly given in the former.

The following instructions were given to the subjects (Kahneman and Tversky 1973):

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100. (p. 241)

This information was presented to a first group of 85 students (low-engineer group). The second group of 86 subjects received the same information, except that the ratio of lawyers and engineers was exactly reversed (high-engineer group). One of the five descriptions goes as follows (Kahneman and Tversky 1973):

Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles. The probability that Jack is one of the 30 engineers in the sample of 100 is ___%. (p. 241)

Because information on resemblance is not given in this study, it only makes sense to compare outcomes between the two groups. For any estimate of $P(EG_H|D)$, the probability that a person is an engineer given the description in the high-engineer group, one can use Bayes' theorem to find the accompanying estimate, $P(EG_L|D)$, for the low-engineer group and vice versa.[17] These derived estimates can subsequently be compared to the actual estimates that the subjects have given. Such a comparison is shown in Figure 2, which is copied from Kahneman and Tversky (1973, p. 242). The curved line shows the correct Bayesian relationship between the probabilities $P(EG_H|D)$ and $P(EG_L|D)$, while the broken line gives the relationship in case no difference in base rate would exist between the high- and the low-engineer group. The five dots represent the median estimates for the five descriptions for the two groups.

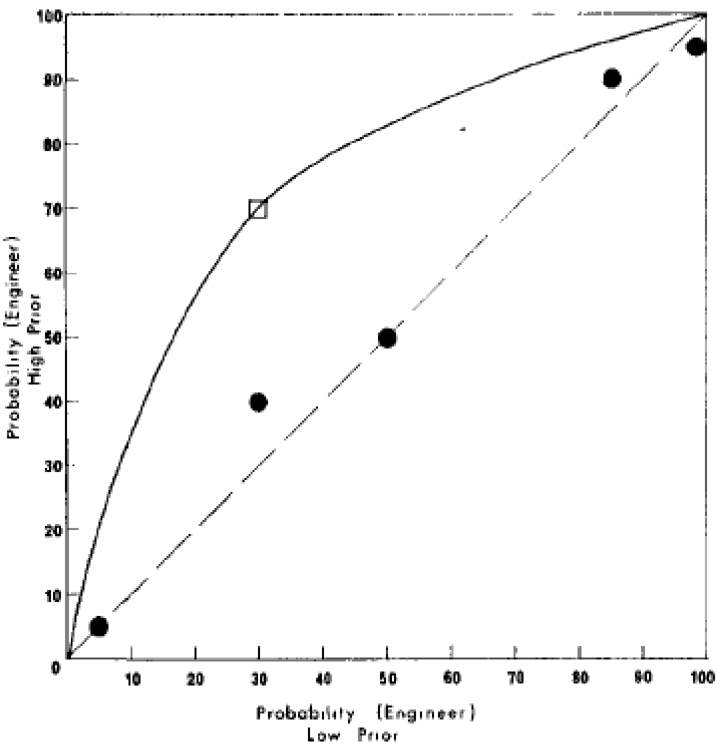

Figure 2

Though Tversky and Kahneman present the results of a statistical test that shows that base rates were not *completely* ignored by the subjects, it is clearly seen from Figure 2 that the subjects' median estimates are closer to the broken line than to the curved line. Tversky and Kahneman argue that these results favor the representativeness hypothesis, as prior probabilities are ignored when individuating information is given. The evidence looms even stronger when one compares the description of 'Dick' to the 'null' description. Kahneman and Tversky (1973) gave Dick the following description:

Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues. (p. 242)

---

[17] To do so, one first calculates the ratio of conditional likelihoods:

$$\frac{P(D|L)}{P(D|EG)} = \frac{(1-P(EG_H|D))*P(EG_H)}{P(EG_H|D)*P(L_H)} \quad [\text{Eq. } 2]$$

With this ratio, one can find

$$P(EG_L|D) = \frac{P(EG_L)}{\left(\frac{P(D|L)}{P(D|EG)}* P(L_L)+P(EG_L)\right)} \quad [\text{Eq. } 3]$$

'Dick' was constructed to contain no useful information, nevertheless the median estimate in both groups that Dick is an engineer was 50%, as seen from the middle dot in Figure 2. This contrasts with the results to the 'null' description, which was given only after the five other descriptions (Tversky and Kahneman 1973):

Suppose now that you are given no information whatsoever about an individual chosen at random from the sample. The probability that this man is one of the 30 engineers in the sample of 100 is ___%. (p. 241)

Subjects correctly estimated this probability to be 30% in the low- and 70% in the high-engineer group, as also shown by the little square in Figure 2. Thus, as Tversky and Kahneman conclude, subjects did attach different values to evidence that is worthless and evidence that is not specific, despite the fact that Bayesian reasoning gives no reason to do so. All in all, Tversky and Kahneman conclude from the Engineer-Lawyer study that base rates are largely ignored due to representativeness.

2.3.2 Concerning the Isomorphism Conditions

How Gigerenzer has evaluated the Engineer-Lawyer study is most clearly seen from his rerun of the Engineer-Lawyer study (Gigerenzer, Hell and Blank 1988, pp. 513-519). Gigerenzer, Hell and Blank criticize the way Tversky and Kahneman deal with the uninformative description of Dick and the condition of random sampling. The other isomorphism conditions are left untouched by Gigerenzer, Hell and Blank's study. I will therefore only shortly discuss these here. This subsection will mainly focus on the assumption of random sampling as dealt with by Tversky and Kahneman and Gigerenzer, Hell and Blank in their Engineer-Lawyer studies. I postpone Gigerenzer's criticism on Tversky and Kahneman's treatment of the case of Dick to section 3.4.2.

Just as in the Tom W. study, the conditional likelihoods are not explicitly given in the Engineer-Lawyer study. Nevertheless one can assume that the subjects had a sufficient intuitive understanding of them, were it only for the fact that the subjects focused so much on representativeness in their estimates. Unlike the Tom W. study, the Engineer-Lawyer study explicitly mentions the relevant base rates. As these are 'exceptionally salient' as Kahneman and Tversky (1973, p. 241) put it, it becomes difficult to criticize them. Because base rates are so clearly put in the final question and because subjects are asked to given estimations in terms of probabilities, it is less tenable to claim that the subjects understood the term 'probable' in a non-mathematical manner.

Gigerenzer, Hell and Blank criticize the random sampling condition in the Engineer-Lawyer study for the reason that the personality descriptions were deliberately constructed to be stereotypical. The description of Jack and the other extreme descriptions, as well as the uninformative description were purposely constructed to be such. This is essentially the same criticism as I already discussed in the context of the Tom W. study. As already explained in section 2.3.2, it is better to ignore base rates in cases in which the evidence is not randomly sampled.

In their study, Gigerenzer, Hell and Blank present two strategies as to better ensure that the assumption of random sampling, which are separately tested. First, four personality descriptions were selected from a sample of real descriptions of engineers and lawyers drawn up by close friends such that the subjects would get the impression that the descriptions were not constructed stereotypes, but randomly picked from the population. Together with the descriptions of Dick and Jack, these were descriptions presented to a group of 48 subjects. The subjects were evenly spread over a high-engineer and a low-engineer group and received exactly the same instructions as in Tversky and Kahneman's study.

As a second strategy, Gigerenzer, Hell and Blank present the information about random sampling not verbally but visually. 10 sheets of papers, with 3 (7) of them marked with an E for engineer and 7 (3) marked with an L for lawyer were put into an urn for the high- (N=25) and low-(N=24) engineer groups respectively. The subject was invited to draw one of the sheets of paper herself from the urn, after which the experimenter show the description to the subject without disclosing the identificatory mark.

Table 5

Base Rate Neglect and Deviations from Bayesian Predictions in Gigerenzer, Hell and Blank's Engineer-Lawyer experiment

	Verbal Random Sampling		**Visual Random Sampling**	
Description	Base Rate Neglect[1]	Bayesian Deviation[2]	Base Rate Neglect[1]	Bayesian Deviation[2]
Thomas	16.7	16.7	23.0	6.6
Peter	15.7	15.8	25.6	6.5
Heinz	2.3	19.9	14.4	10.6
Jack	9.9	16.1	15.1	7.8
Klaus	13.0	11.4	8.1	15.5
Dick	1.2	30.3	17.8	13.8
Null	40.8	-3.8	41.3	-1.4
Average (excl. Null)	9.8	18.4	17.3	10.1

[1] The differences between the means in the 70% and 30% base rate conditions
[2] The differences between the means of the low-engineer group and the Bayesian predictions for the low-engineer group derived from the high-engineer group

Results to Gigerenzer, Hell and Blank's version of the Engineer-Lawyer study are shown in Table 5. The 'Base Rate Neglect' columns show the difference between the estimates in the high- and the low-engineer groups. The closer to zero these results are, the more base-rate are neglected. For the 'Bayesian Deviation' columns, first the mean estimates of the high-engineer group were used to calculate the estimates these subjects would have with the base rates of the low-engineer group, holding the conditional likelihood constant.[18] Subsequently, these estimates were subtracted from the mean replies of the low-engineer group. The closer to zero, the more these results are in line with Bayesian reasoning.

From the results, Gigerenzer, Hell and Blank first conclude that base rate neglect is lower with verbal random sampling than it was in the original study of Tversky and Kahneman. As Gigerenzer, Hell and Blank suggest, this might have to do with the fact that the descriptions were less artificial, such that subjects put more trust in the information on random sampling. Alternatively though, base rates may simply have more impact on less stereotypical descriptions for the reason that subjects are less likely to depend on the representativeness heuristic as a shortcut procedure for such cases. The shortcut procedure can be explained as follows. As Figure 2 shows, the differences in Bayesian estimates for different base rates are smaller when descriptions are more stereotypical. A subject may therefore not to go into the effort of discounting resemblance with the prior probabilities, if the latter only have a small effect. If the subject uses this shortcut procedure for one or two personality

[18] Essentially using equation 3 and 4.

descriptions, it might become some sort of 'anchor' for subsequent personality descriptions. I will discuss this issue more substantially in section 3.4.2, when I discuss the case of Dick.

Furthermore, Gigerenzer, Hell and Blank conclude that Bayesian reasoning improves with visual random sampling. They use the following index b of base rate use to express the degree of base rate use numerically:

$$b = \frac{Base\ Rate\ Neglect}{Base\ Rate\ Neglect + Bayesian\ Deviation}\ [Eq. 7]$$

Low values of b show that the subjects' mean judgments are showing more base-rate neglect, while values closer to 1 indicate increasingly correct Bayesian reasoning. For the verbal random sampling groups, b=0.35, for the visual random sampling group, b=0.63. Thus, Bayesian reasoning significantly improved with visual random sampling, though the latter did not eradicate neglect of base rates.

2.3.3 Tversky and Kahneman's reply

In this subsection I will discuss Tversky and Kahneman's possible replies to Gigerenzer's two point of criticism towards the Engineer-Lawyer study. The most striking issue though about the Engineer-Lawyer study is that Tversky and Kahneman and Gigerenzer have agreed on most details of the Engineer-Lawyer experiment. Of all possible lines of criticism, Gigerenzer has only focused on random sampling. By conducting his own version of the Engineer-Lawyer experiment, he has implicitly argued that all other conditions are satisfied. In that sense, Gigerenzer and Tversky and Kahneman actually work quite close to each other.

Last paragraph's conclusion is surprising if one considers the drastic interpretation that Gigerenzer has attached to the results of Gigerenzer, Hell and Blank (1988) in his paper 'How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases"'. In this paper Gigerenzer argues that visual random sampling made base-rate neglect disappear (Gigerenzer 1991A, p. 98). Kahneman and Tversky (1996, p. 584) have criticized this exaggerated interpretation of Gigerenzer, Hell and Blank (1988), arguing that Gigerenzer's 'announcement about the disappearance of base-rate neglect is premature'. In line with studies in the heuristics and biases program, Gigerenzer, Hell and Blank (1988) only show underweight of base rates.

One should be aware that Kahneman and Tversky (1996) only criticized Gigerenzer's (1991A) unfounded interpretation of Gigerenzer, Hell and Blank (1988). Towards Gigerenzer, Hell and Blank's findings on visual random sampling, Kahneman and Frederick (2002, p. 69) are actually quite sympathetic. They argue that is one way to manipulate subject's attention towards the neglected variable, just like one can add vertical lines in the Müller-Lyer illusion. Obviously, when subjects are shown that 10 sheets of paper are put into an urn, after which they can themselves draw one of the sheets, their attention is focused on the Bayesian structure of the problem.

With the visual procedure, Gigerenzer, Hell and Blank have put more attention to the internal representation of the subjects, who may come to think that they are fooled about the random sampling information given by the experimenter. At the same time visual random sampling is much closer to 'Socratic instruction' than verbal random sampling. As in the case of Tom W., one can quibble endlessly about the question how much guidance a subjects is allowed to receive, such that an experiment qualifies as a fair test of intuitions. I would like to stress here again that there are no clear-cut rules on this matter. What we can safely conclude though from the various Engineer-Lawyer studies is that different results for a change in the presentation of random sampling show how subjects perform for different levels of guidance. In that sense, the results found by Gigerenzer, Hell and Blank (1988) and Kahneman and Tversky (1973) are simply enriching one another.

Tversky and Kahneman have not replied to Gigerenzer's criticism on the use of fabricated personality descriptions. I expect them not to have much difficulty towards Gigerenzer, Hell and Blank's approach. Their study still shows neglect of base rates, just like the original study. The result is maybe less outspoken, but the results are nevertheless in line with the literature as Kahneman and Tversky (1996, p. 584) notice. One might even expect that the more representative a personality descriptions is of a certain occupation, the more likely people are to give a judgment by representativeness[19], for example because of the shortcut procedure introduced in last subsection. As real personality descriptions are less stereotypical and therefore less representative of an occupation, the probability estimates will be more tempered. The use of real personality descriptions is in that sense not hostile to the heuristics and biases program, as it shows to what extent subjects depend on information of representativeness when making probability estimates. Moreover, Gigerenzer's criticism towards deliberately constructed information has sparked studies in the heuristic and biases program, such as the one presented by Slovic, Monahan, and MacGregor (2000) to use actual instead of fake cases.

Despite the rhetoric by Gigerenzer (1991A), Tversky and Kahneman and Gigerenzer have acted in close harmony with their respective Engineer-Lawyer studies. The only aspect on which Gigerenzer has disagreed with Tversky and Kahneman, namely random sampling, has foremost inspired him to conduct a new study. This new study has in the end been welcomed by the heuristics and biases program and has not sparked fundamental controversy.

2.4 The Taxicab Problem

The reason why Tversky and Kahneman came to conduct experiments involving the use of Bayes' theorem was through the evidence that they found in other types of studies on representativeness. Bayesian problems, popular as they were in experimental studies in the 1960's, presented themselves as another opportunity to test the hypothesis that subjects predict by using the representativeness heuristic. Indeed, Tversky and Kahneman found considerable evidence that representativeness was the driving force in problems that can be solved by Bayesian reasoning. But what if a sample is not (obviously) representative of the population; does that make people reason flawlessly?

In this subsection, I will not only introduce the Taxicab Problem and Gigerenzer's criticism towards it, but I will also present a few other base-rate neglect studies. My goal thereby is to focus on the care that researchers working in the heuristics and biases program gave to the context of problems. Besides Tversky and Kahneman's paper, the work of Lyon and Slovic (1976) and Bar-Hillel (1980) are also taken to be part of the heuristics and biases program. As Tversky and Kahneman (1982B, p. 154) have been supportive towards both Lyon and Slovic (1976) and Bar-Hillel (1980), the reader is safe to assume that these authors are much in agreement with one another.

Through this care for context, heuristics and biases researchers have been able to improve the design of their experiments, along the same lines as Gigerenzer has proposed his improvements. Moreover, the focus on the context of judgmental problems invalidates Gigerenzer's claim that the heuristics and biases program has often ignored the influence of context, to which I will return in section 3.5.5. As part of these aims, I discuss multiple versions of Tversky and Kahneman's Taxicab Problem in this section, as well different versions of Hammerton's studies on the use of base rates.

[19] The reader who is puzzled about this relation, may want to consult section 3.1 first, which introduces the difference between judgments of and judgments by representativeness.

2.4.1 The Early Taxicab Problem

In 1972, Tversky and Kahneman designed an experimental study, the Taxicab Problem, which addresses the question whether people still neglect base rates when judgmental problems are not designed to contain representative descriptions. Peculiarly though, it was not at all the reason for Tversky and Kahneman to conduct this study. They reasoned that they had only found evidence for 'the discarding of prior probabili[ti]es' when subjects are given psychological evidence (Kahneman and Tversky 1972A, p. 13). The Taxicab Problem was designed to confirm that phenomenon in a different context. The original problem as formulated by Kahneman and Tversky (1972A) follows:

Two cab companies, the Blue and the Green, operate in a given city. 85% of the cabs in the city are Blue, and the remaining 15% are Green. A cab was involved in a hit and run accident at night.
A witness identified the cab as a Green cab. The court tested his ability to distinguish a Blue cab from a Green cab at night, and found that he was able to make correct identifications in 4 out of 5 tries. (p. 13)

Presented with this information, the subjects were asked to assess the likelihood that a Blue cab rather than a Green cab was involved in the accident (Kahneman and Tversky 1972A). In contrast with Tom W. and the Engineer-Lawyer studies, the Taxicab Problem does not present descriptions that are stereotypical to some population or at least resemble it to a relatively high extent. In more formal terms, these first experiments present a class *M*, such as engineers or students and some sample *X*, which is a subset of *M* (Kahneman and Tversky 1982B: 86). Defined as such, representativeness has no bearing on this problem. The median estimated probability in the pilot study conducted by Tversky and Kahneman was 0.20, whereas Bayes' theorem gives that the probability of a blue cab (*B*) given the witness report (*W*) is:

$$P\big((B|W)\big) = \frac{.2 * .85}{(.85 * .2 + 8 * .15)} \approx 0.41 \text{ [Eq. 4]}$$

The subjects' estimates suggest that base-rate neglect is not restricted to the representativeness heuristic.

2.4.2 The Hammerton Studies

In 1973, Hammerton published a study with results in line with Tversky and Kahneman's pilot study. His first study presents the subjects with a medical screening test. Subjects were given the following instructions (Hammerton 1973):

1. A device has been invented for screening a population for a disease known as psylicrapitis.
2. The device is a very good one, but not perfect.
3. If someone is a sufferer, there is a 90% chance that he will be recorded positively.
4. If he is not a sufferer, there is still a 1% chance that he will be recorded positively.
5. Roughly 1% of the population has the disease.
6. Mr. Smith has been tested, and the result is positive. The chance that he is in fact a sufferer is: (p. 252)

The median estimate of the group of 10 subjects to this question was 0.85, while the correct Bayesian answer is approximately 0.48. Hammerton argues that the reason why subjects gave such extreme probabilities is due to the high priors that they had on medical tests. People expect these to be infallible, which gives the estimated probabilities an upward prejudice. To test for this hypothesis, Hammerton presented the problem to another group of 20 subjects, leaving the structure untouched, but with a change in content. Instead of a medical test, the problem discusses that 'a device has been invented for screening engine parts for internal cracks' (Hammerton 1973, p.254). In line with Hammerton's expectation, the median estimated probability was 0.6, showing a smaller deviation from the correct Bayesian result than the original study.

Both Hammerton's and Tversky and Kahneman's study found that the subjects' estimates were dominated by the stated accuracy of the individuating data. However, Hammerton did not found this domination for a case in which subjects allegedly did not have rigid priors, whereas Tversky and Kahneman did. This was a reason for Lyon and Slovic (1976) to delve into this topic further, mainly by conducting improved versions of the studies just described. For the Taxicab Problem, Lyon and Slovic specified further that the accuracy of the witness in making correct identifications was 0.8 *regardless of color*. Furthermore, they criticized the original 'Engine Crack' problem for using the phrase 'the result is positive'. Outside a medical context, the meaning of 'a positive result' might be ambiguous, as it may both indicate that the engine part is OK, or that it is broken. Slovic and Lyon replaced this information with phrases that are more descriptive. For all versions of the experiments, the median estimates show large deviations from the Bayesian answers, although some are more accurate than others.

A criticism that Gigerenzer (1991A, p. 96) has posed against the Harvard Medical School test as conducted by Casscells, Schoenberger and Grayboys (1978, p. 999) is that medical screening tests do not presuppose random sampling. In practice, doctors only see people who have health problems. This raises the chances that they suffer from some specific disease compared to healthy people. Thus, under normal circumstances no random sampling is involved. Neither the Harvard Medical School test, nor Hammerton's study suggest a change in these conditions. Therefore, one should not be surprised to find results that deviate from the estimate generate with Bayes' theorem, which does presuppose random sampling. Although this criticism by Gigerenzer certainly makes some sense, this criticism has already been taken up by Hammerton and Slovic and Lyon as they also conducted experiments with the Engine Crack problem, for which random sampling appears to be the default condition.

Hammerton's studies are still not without problems though, as Gigerenzer might come to argue. Specifically, the second instruction in Hammerton's problems is misleading. The instruction that 'the device is a very good one, but not perfect' only applies to $P(P|H)$ and $P(P|\neg H)$. With *P* denoting probability of a positive result and *H (*$\neg H$*)*, the hypothesis that the patient is indeed sick or the engine indeed broken is (not) true. Subjects might come to think that the instruction applies to $P(H|P)$as well. The alleged high quality of the device might have very well induced subjects to underrate the possibility of false negatives ($P(P|\neg H) * P(\neg H)$). Indeed, one would not sensibly expect a 'very good' device to be that uninformative as the Bayesian estimate of only .48 predicts it to be. In other words, one could argue that the condition of posterior probability is neither safeguarded in the original Hammerton's studies, nor in Slovic and Lyon's versions.

2.4.3 The Later Taxicab Problem

Tversky and Kahneman returned to the Taxicab Problem in the article 'Causal Schemas in Judgments under Uncertainty', first published in 1977.[20] The claim of this paper is that the psychological impact of pieces of evidence in judgment under uncertainty depends on their role in a causal schema. In a causal schema, a course of events is organized into cause-effect relations. People rely on causal

[20] Other versions were published in 1977(b) and 1980. Most authors refer to the 1980 version, which is partly reprinted in Kahneman, Slovic and Tversky (1982).

schemas to achieve a coherent interpretation of the events that surround them. From this, Tversky and Kahneman hypothesize that people will ignore base rates, unless they can be given a causal interpretation, regardless of representativeness. The Taxicab Problem is invoked to provide evidence for this hypothesis.

Tversky and Kahneman distinguish three different ways in which people classify the relation between evidence or data *D* and some event *X*. If people believe that *D* is the cause of the occurrence or non-occurrence of *X*, *D* is referred to as a *causal* datum. D is a *diagnostical* datum when *X* is seen as a potential cause of *D*. Finally, when the subject fails to see indirect or direct causal links, Tversky and Kahneman treat it as *incidental*.

Tversky and Kahneman (1977B, p. 1-2) claim that incidental 'causal data has a greater impact than diagnostic data of equal informativeness, and that incidental data are given little or no weight, in the presence of causal or diagnostic data'. Allegedly, people find it easier to interpret causal data as it resembles the natural course of a causal schema, in which causes are followed by consequences. If the order is reversed, such as for diagnostical data, people find it more troublesome to fit the evidence into a causal schema. As a result, people give more weight to causal data than to diagnostic data. Likewise, incidental data will be given little or no weight when presented together with causal or diagnostic data. Tversky and Kahneman argue that this treatment of data is not in line with the normative treatment of probability (1982A, p. 118). Only the informativeness of *D* should matter for it impact, the relation to which it stands to X is irrelevant.

Through a series of experiments Tversky and Kahneman tested their claim that causal data are given greater importance. Subjects were asked to compare two conditional probabilities, $P(X|Y)$and $P(Y|X)$, with $X = Y$, such that $P(X|Y) = P(Y|X)$. *X* was framed such that it was seen as a natural cause of *Y*. As an illustration, consider the following study, which is taken from Tversky and Kahneman (1982A):

In a survey of high-school seniors in a city, the height of boys was compared to the height of their fathers. In which prediction would you have greater confidence?
(a) The prediction of the father's height from the son's height. (N = 23)
(b) The prediction of the son's height from the father's height. (N -=53)
(-) Equal confidence. (N = 76) (p. 119)

In accordance with Tversky and Kahneman's hypothesis, subjects regarded *X* (the father's height) to be stronger evidence for *Y* (the son's height) than vice versa. With this newly gained knowledge, Tversky and Kahneman (1977A) hypothesize that 'in the presence of specific evidence, base-rate data will be essentially ignored, except when they can be incorporated in a causal model which applies to the case under consideration' (p. 174). Tversky and Kahneman test this claim through an experimental study of the Taxicab Problem. I here present the version presented in Tversky and Kahneman (1982B), as the older versions do not sufficiently specify the conditional likelihoods.

A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
(a) 85% of the cabs in the city are Green and 15% are Blue.
(b) a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green? (pp. 156-157)

When Bayes' theorem is used for this problem, one finds that the probability that the car was blue (*B*) given the witness report (*W*) is:

$$P\big((B|W)\big) = \frac{.8 * .15}{(.8 * .15 + .85 * .2)} \approx 0.41 \text{ [Eq. 5]}$$

Yet, the median estimate was found to be 0.8, indicating that base rates were completely ignored by the subjects. As the ignorance of base rate information in the Taxicab Problem is not a result of the representativeness heuristic, Tversky and Kahneman conclude that the neglect of base rates is a very general phenomenon.

In a different version of the experiment, Tversky and Kahneman replaced (a) with (a'), to give the base rates a causal interpretation:

(a') Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs.

The median estimate given was 0.6, showing only moderate base rate neglect. Given the estimate to the formulations of the Taxicab Problem presented here, Tversky and Kahneman (1977A,) conclude that 'base-rate information which is not linked to a causal schema is ignored in the presence of causally relevant evidence' (p. 178). According to Tversky and Kahneman intuitive reasoning thereby significantly deviates from the normative theory of evidence (1977B, p. 3-17).

2.4.4 Gigerenzer's criticism on the Later Taxicab Problem

Gigerenzer and Murray (1987) do not agree with Tversky and Kahneman's conclusion. They argues that the subjects' reasoning was not flawed at all, but that the experimenters instead were mistaken! Although Gigerenzer and Murray do not put it that way, the reason for their criticism is that the assumption of equal informativeness between diagnostic and causal data does not hold. Contrast to the father-son-height problem discussed in last subsection, where the relation between data and the event did not matter for the informativeness of the data, one could argue that this is different for the Taxicab Problem. Specifically, the isomorphism conditions for base rates and random sampling are not fulfilled because of the absence of a (sufficiently specified) causal relation.

According to Gigerenzer and Murray (1987, p. 161) the base rates are underspecified in the Taxicab Problem. Put differently, there is no reason to assume that the cab involved in the accident was randomly sampled from the given population. Knowing that 85% of the cabs in the city are green or that 85% of the green cabs are involved in accidents, does not say anything about the fraction of green cabs that is involved in *hit-and-run* accidents *at night*. Contrary to what Tversky and Kahneman suggest, the base rates do differ in informativeness. The fact that 85% of the green cabs are involved in accidents simply gives more information than the fact that 85% of the cabs in the city are green. If subjects think along these lines, they are right to neglect or at least give less value to base rates!

One should recollect that Bayesian reasoning does not force a subject to use the base rates given by the experimenter. As Gigerenzer (1991A, p. 88) has emphasized, any information might be used as an input for the prior probabilities. This is not restricted to base-rate information, but may include any belief that a subject has on some matter. Indeed in the Taxicab Problem, someone may choose to incorporate other information than base rates (only) in the prior probabilities as the base rates do not or only weakly apply to the specific problem. This has been a reason for Gigerenzer carefully distinguishes between base rates and prior probabilities. Base rates may be, but are not necessarily the same as the prior probabilities that enter Bayes' theorem.

The point made by Gigerenzer and Murray (1987) concerning equal informativeness may explain the peculiar results to the Intercom problem, a special version of the Taxicab Problem, which was first presented by Bar-Hillel (1980). Bar-Hillel thought that Tversky and Kahneman's hypothesis on causal models was too narrow and hypothesized that subjects ignore base-rate information whenever they feel it is not *relevant* to the judgment they are asked to make. Thus information that is judged to be more relevant will gain more weight in judgment. Thus, to make sure that base rates and indicant information are given equal weight, one should induce people to give equal relevance to both pieces of information. Bar-Hillel does so by lowering the relevance of the indicant information, which loses its causal interpretation in the Intercom problem. [21]

In the Intercom problem, subjects are told that 85% of the cabs are Blue and 15% are Green. Subjects are informed that the witness heard the sound of an intercom coming through the cab window. Police investigations show that intercoms are installed in 80% of the Green cabs, and in 20% of the Blue cabs. Subject gave a median estimate of .48 for the probability that a Green cab was involved in the accident, which is only slightly above the Bayesian estimate of .41. Estimates were highly dispersed though, as shown in Figure 3[22], which gives the distribution of responses to the Intercom Problem. Bar-Hillel has not attempted to explain these results.

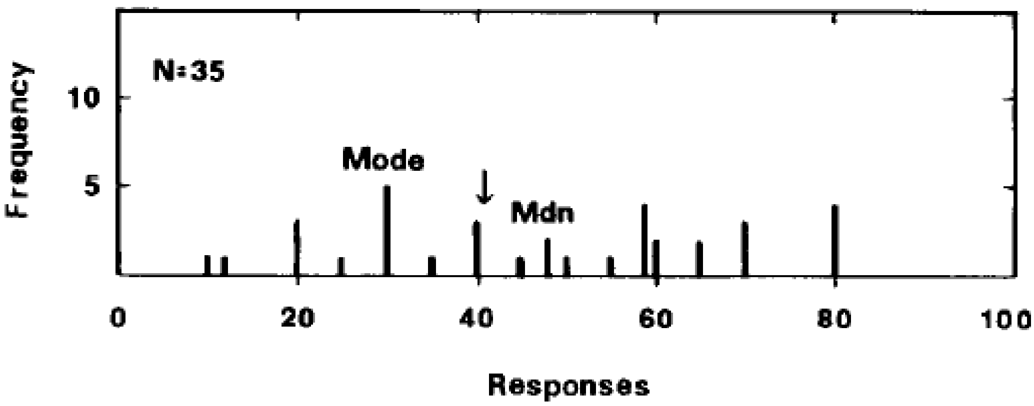

Figure 3

Interestingly, both the base rates and the indicant information does not tell one which cab is more likely to be involved in hit-and-run accidents at night as Gigerenzer and Murray (1987) comment one would like to know. It is therefore not surprising that subjects give nearly equal weights to both pieces of information. A further complicating fact is that subjects are induced to believe that the indicant information is relevant, as it is a result of police investigations. Why would the police conduct an investigation when this does not increase relevance? Thus, subjects may have come to take the indicant information more seriously than the base rate information for this reason, which may explain why the estimates of the subjects show an upward bias. One can conclude from this that Gigerenzer's criticism seems to be on the right track, if only in giving an explanation of what goes on.

[21] Tversky and Kahneman (1982B) argue that Bar-Hillel's results show that 'base-rate data are combined with other evidence either when the former have a causal interpretation or when the latter are no more specific than the base rate' (p. 158).

[22] Copied from Bar-Hillel (1980, p. 228)

2.4.5 Tversky and Kahneman's reply

Tversky and Kahneman have not responded to Gigerenzer's claim that the implicit assumption of equal informativeness does not hold in the different versions of the Taxicab Problem. Bar-Hillel (1980, pp. 217-218) though, has discussed the informativeness of base rates and indicant information, just like Gigerenzer and Murray (1987) did on the Taxicab Problem. She argues that when information is given about both a set and a subset, the information on the latter is more relevant for making judgments about members of the subset. In line with Gigerenzer, she claims that it is justified that bases rates are superseded by specific information in such cases. Surprisingly though, Bar-Hillel does not think that the base rates in the Taxicab Problem should be ignored, despite admitting that the base rates in the original problem are far more general than the witness' testimony. However, Bar-Hillel argues, because the testimony is not perfectly reliable, it should not supersede the base rate.

I do not think that Gigerenzer would find this criticism much convincing. The absolutely measured unreliability of the testimony does not allow us to draw conclusions about its usefulness in relation to the base rates. In case one is unsure whether the base rates apply to the case at hand, there is ample reason to ignore or diminish the impact of base rates and to solely focus on the more specific information instead, despite of the latter's unreliability. In conclusion, I see no clear reason why researchers in het heuristics and biases program should not accept Gigerenzer's criticism on the Taxicab Problem.

Although Gigerenzer's criticism give reason to reconsider the results to the Taxicab Problem, this may not hurt Tversky and Kahneman's claims about the importance of causal schemata, as these do not depend on the Taxicab Problem. Another study, the Suicide Problem, has also been drawn on to support these claims and this study has been left untouched by Gigerenzer. In the Suicide Problem, which Tversky and Kahneman (1977a), the base rates are specific enough to be used for judgment:

Consider the following hypotheses regarding suicide. In a population of young adults, 80% of the individuals are married and 20% are single. The percentage of deaths by suicide (relative to all other forms of non-natural deaths) is 15% among single individuals and 5% among married individuals.
Question: What is the probability that an individual, selected at random from those that had committed suicide, was single? (p. 177)

The median estimate of the 73 subjects was .75, with the correct Bayesian answer being 3/7. With the following design of the problem as presented by Tversky and Kahneman (1977a), which allows for a causal interpretation of the base rate data, the median estimate changed to 0.5 however:

Consider the following hypotheses regarding suicide. In a population of adolescents, 80% of suicide attempts are made by girls, and 20% by boys. The percentage of suicide attempts that result in death is 15% among boys and 5% among girls.
Question: What is the probability than an adolescent, selected at random from those who had died by suicide, was a girl? (p. 178)

The major conclusion of this research is that the use or neglect of consensus information in individual prediction depends critically on the interpretation of that information. Notice that the Suicide Problem does not run into the same troubles with the isomorphism conditions as the experiments discussed thus far. Representativeness does not play the role in this problem as it did in the Tom W. and the

Engineer-Lawyer studies, such that no problems are encountered with the interpretation of the semantic conversion of the posterior probability. Likewise, the case is not deliberatively constructed to be stereotypical, so no problems with the interpretation of random selection in this problem. Moreover, unlike the Taxicab Problem, the base rates do apply to the question asked.

What the researchers in the heuristics and biases program have failed to acknowledge is that the role of perceived relevance is not only of empirical interest, but also of normative interest. Gigerenzer and Murray (1987) have argued that a subject may credibly argue that base rates do not apply. Thus, even though researchers in the heuristics and biases program have not ignored the context of a problem, they do have failed to sufficiently take its normative repercussions into account for the Taxicab Problem. Nevertheless, we have seen that Gigerenzer's criticism does not hurt all studies of the heuristics and biases program. Especially, the Suicide Problem is left unimpaired by Gigerenzer.

A second aim of this section has been to show that researchers in the heuristics and biases program have made improvements to the judgmental problems along the same lines as Gigerenzer. Specifically, Lyon and Slovic have increased the clarity of the conditional likelihoods in the Hammerton studies and the Taxicab Problem. One may of course comment on their treatment of these studies that they still can be criticized from Gigerenzer's perspective as I have shown in section 2.5.2. These comments do not have to be regarded as destructive though, but can be seen as elaboration on earlier improvements to the design of these studies.

Now that I have discussed the most important base-rate neglect studies of the heuristics and biases program, I will wrap up the discussion on the isomorphism conditions in the next subsection.

2.5 Conclusion

Both Gigerenzer and Tversky and Kahneman aim 'to understand the cognitive processes that produce both valid and invalid judgments'. Gigerenzer has criticized Tversky and Kahneman's delineation between valid and invalid judgment though, as we have seen in this section. Gigerenzer argues that such a delineation can only be made when Bayes' theorem is properly applied. On its own, this is a rather uncontroversial claim. Basically everyone would agree that Bayes' theorem only applies when certain conditions are satisfied. One may disagree though on what these conditions are and to what extent they are to be satisfied.

I have shown that Gigerenzer's critique can be summarized under four headings namely 'base rates', 'conditional likelihoods', 'posterior probability' and 'random sampling'. I have used this framework to cover Gigerenzer's points of criticism towards the studies performed within the heuristics and biases program. On top of that, I also used it to formulate criticism towards these studies that Gigerenzer has not actually cast, but which is in line with his other work.

Taking a look at a Table 3 again, one can see that the 'Vs' clearly outnumber the 'Xs' and the '?s'. Thus, there are more points of agreement between Gigerenzer and Tversky and Kahneman than points of disagreement. This shows that Gigerenzer has been quite supportive towards the studies of base-rate neglect heuristics and biases program. Instead of disqualifying this branch of research altogether, Gigerenzer has joined the Bayesian framework by giving specific points of criticism and foremost by conducting base-rate neglect studies himself. This has been most outspoken in Gigerenzer, Hell and Blank's rerun of Tversky and Kahneman's Engineer-Lawyer study. I have argued that this study has been constructive from the point of view of the researchers in the heuristics and biases program.

Although Gigerenzer has accepted the use of Bayes' theorem in tests of statistical intuitions as a general framework, still many issues between him and Tversky and Kahneman remain unresolved, as one can see from the '?s' in Table 3. I have argued that the presence of so many question marks is a result of the subjective theory of probability Tversky and Kahneman subscribe to. As a result of this theory, it is the subject that in the end has to judge whether her probability estimates are consistent with her entire web of beliefs. Thus, it is also the subject who has to agree that Bayes' theorem can be applied to a judgmental problem in the manner envisaged by the experimenter.

The implication of the subjective theory is that any criticism that one could have on the applicability of Bayes' theorem, should be taken seriously. In principle, if there is a reason to think that the subject has understood the question differently than the experimenter had pictured, then this provides a reason to doubt the validity of a test. As has been discussed in detail though, a subject's interpretation of a problem is notoriously hard to retrieve. Consequently, it is difficult to assess how widespread among the population a certain interpretation of a problem is. Notice the complicating factor that people may not think uniformly about the interpretation of a problem. As a result it is hard to come up with clear cut rules that specify what a fair test of intuitions is. Of course one could take as many precautions as possible to make sure that subjects interpret a problem along the desired lines of the experimenter, but such an approach runs into the problem that an experiment may no longer be a test of intuitions but an exercise of Socratic instruction instead. It appears that one can discuss endlessly about whether a specific problem constitutes a fair test of intuitions, which explains the '?s' in Table 3.

As emphasized by Gigerenzer (1996), he and researchers in the heuristics and biases program alike both aim 'to under-stand the cognitive processes that produce both valid and invalid judgments' (p. 592). We have seen that it is difficult to draw a line between valid and invalid judgments. Consequently, it is understandable that Gigerenzer and Tversky and Kahneman may choose different sides when it comes to some of the details. This does not exclude though that Gigerenzer and Tversky and Kahneman can work side by side when it comes to understanding the underlying processes. Not surprisingly, Tversky and Kahneman and Gigerenzer have frequently relied on one another's experimental studies. Thus, despite the harsh rhetoric of for example Gigerenzer (1991A), but also Kahneman and Tversky (1996), it appears that the discussion between the two camps concerning the applicability of Bayes' theorem is much more nuanced.

Before drawing my final conclusion, I will first give a summary of section 2.2-2.4. In section 2.2 I have argued that from the Gigerenzerian framework, one can criticize the conditions of 'random sampling' and 'posterior probability' in the Tom W. study. I have shown that Gigerenzer has been simply wrong in claiming that in the original Tom W. studies random sampling is not mentioned at all. This should be regarded as a missed opportunity for Gigerenzer. He could still have criticized this study for using a deliberately constructed description, like he has done for the Engineer-Lawyer study. By invalidating the Tom W. study for only one incorrect reason, Gigerenzer has neglected the opportunity to criticize the Tom W. experiment on other grounds, such as on posterior probability.
I have argued that this issue can be solved by using frequency formats instead of probability formats, as advocated by Gigerenzer and Hoffrage (1995). Finally, I have argued that the criticism on posterior probability does not present a knock-down argument against Tversky and Kahneman, such that the latter are not forced to change their position.

The Engineer-Lawyer study has been criticized by Gigerenzer for not assuring the condition of random sampling. He argues that the personality descriptions presented by Tversky and Kahneman are deliberately constructed to be stereotypical, such that subjects had a reason to doubt the assumption of random sampling. Moreover, he has claimed that informing subject verbally about random sampling is not sufficient in the Engineer-Lawyer study. Instead, random sampling should occur visually, on the spot. These criticisms have led Gigerenzer to conduct his own experiments with real descriptions and visual random sampling, of which the results are acknowledged by Tversky and Kahneman.

Finally, Gigerenzer has criticized the different versions of the Taxicab Problem for not giving base rates that are equally informative. Specifically, Gigerenzer has argued that 'causal' base rates are more informative than the 'incidental' base rates. I have argued that researchers in the heuristics and biases program should accept this normative criticism. From an explanatory point of view, no disagreements exist between Gigerenzer and the heuristics and biases program. Both have emphasized that (perceived) relevance or more specifically the causal role of base rates is decisive for the subjects' estimates. I have furthermore shown in section 2.4 that Gigerenzer's framework can also be applied to other base-rate neglect studies, such that previously unmentioned weaknesses and accompanying improvements become apparent.

The summary of section 2.2-2.4 again show that Gigerenzer has been constructive towards the heuristics and biases progam's base-rate neglect studies, which is again most clearly seen from the Engineer-Lawyer study. At the same time, Gigerenzer has at times fallen short of applying his own framework. In the case of the Tom W. study, Gigerenzer could have more careful. Because of his carelessness concerning random sampling though, he has unnecessarily distanced himself from Tversky and Kahneman. A rerun of the Tom W. study with real instead of fabricated personality descriptions and with a bit more careful wording would have already done the deal as my analysis shows.

In conclusion, I would argue that Gigerenzer has been constructive towards the heuristics and biases program's base-rate neglect, although it must be added that he could have been more constructive without much additional effort. At least, the two parties are not as divided as the rhetoric sometimes suggests, as they agree on most issues. It was not for no reason that I had to focus on so much detail: only by zooming in on specific cases I could demonstrate on what issues Gigerenzer and Tversky and Kahneman agree and disagree.

Finally, I will now discuss how my conclusions relate to the literature reviewed in section 1.2. Firstly, my thesis nuances the view put forward by Samuels, Stich and Bishop (2002). They have pointed out though that the issue of the applicability of Bayes' theorem do form real disputes between the two camps. In contrast, I have shown that this discussion has been mostly constructive.

Secondly, my conclusion seems to be very much in line with Vranas (2000, 2001). He argued that Gigerenzer's problems are not so much with the validity of abstract probabilistic norms, but with statistical norms that apply to specific content. My detailed analysis of Gigerenzer's criticism towards base-rate neglect studies indeed shows that much of his concern goes to this issue. The confusion in the debate between Vranas and Gigerenzer might be explained by Gigerenzer's claim that the use of statistical tools makes the accompanying probabilistic norms often unreflectively translate into statistical norms. I will introduce this claim in section 3.5, where I will show that this claim is overblown towards Tversky and Kahneman's use of Bayes' theorem, as well as towards their use of the representativeness heuristic, which according to Gigerenzer stems from Bayes' theorem.

Thirdly, my account is not incompatible with that of Jullien and Vallois (2014). They have argued that the difference in probabilistic theory adhered to, drives the different choices that both camps make in relating context and norms. Gigerenzer, who is a frequentist, reasons from context to norms, while Tversky and Kahneman, who are subjectivists, reason from norm to context according to Jullien and Vallois.

I would argue though that from a practical point of view, the dichotomy sketched by Jullien and Vallois does not have many implications. Both have reason to study how the context of a problem, either as being perceived by the subject or as being objectively given, affects the way a person interprets a problem. From what angle they do so, does not really matter in practice.

3. On the Representativeness Heuristic

3.1 Judgments of and by Representativeness

As discussed in section 1.4.1, Tversky and Kahneman were initially only interested in the prevalence of the representativeness heuristic in different contexts. What they did not realize though, is that they had been testing for different forms of representativeness. Only in 1982, with the publication of 'Judgments of and by Representativeness' Tversky and Kahneman came to explicitly distinguish the two forms of representativeness introduced in the title of the article. In this subsection I will explain how judgments of and judgments by representativeness relate. Judgments of representativeness concern the nature of the representativeness relation, while judgments by representativeness concern the use of representativeness for judgments and prediction. As we will see, a thorough understanding of these two forms of representativeness is key in evaluating Gigerenzer's criticism on the representativeness heuristic.

Let me start off with providing Tversky and Kahneman's rather abstract definitions of judgments of and by representativeness. Subsequently, I will try to clarify both concepts by referring to the many studies already discussed in this thesis. Kahneman and Tversky (1982B) present two hypotheses regarding representativeness, of which claim that they are conceptually independent. The first hypothesis, which concerns judgments *of* representativeness, claims that 'people expect samples to be highly similar to their parent population and also to represent the randomness of the sampling population' (Kahneman and Tversky 1982B, p. 84). Representativeness is now defined as 'a relation between a process or a model, *M*, and some instance or event, *X*, associated with that model' (Kahneman and Tversky 1982B, p. 85).

Kahneman and Tversky (1982B, pp. 85-87) distinguish four basic cases of judgments of representativeness. The fourth case is remarkable, as it denotes a causal relation. The causal schemata which were separated from the representativeness heuristic in Tversky and Kahneman (1977AandB, 1980) are now put under the header of representativeness, such that the results to the Taxicab experiments should be interpreted as driven by the representativeness heuristic. The four basic cases follow below:

1. M is a class and X is a value of a variable defined in this class.
Thus, M might be a group of students and X the representative value of their age, weight, length, income, grades etc.
2. M is a class and X is an instance of that class.
So, if M is the class of birds, a robin is found to be more representative of that class than a chicken. Case 2 is a special case of case 3.
3. M is a class and X is a subset of M
Within the class of students, the subset of Psychology students are more representative than engineering studies.
4. M is a (causal) system and X is a (possible) consequence.
People believe that some causal system M, say rain, causes floods as a consequence.

The second hypothesis, which concerns judgments *by* representativeness, claims that 'people often rely on representativeness as a heuristic for judgment and prediction' (Kahneman and Tversky 1982B, p. 84). This means that people calculate the probability of an uncertain event to the degree that it is '(i) similar in essential properties to its parent population and (ii) reflects the salient features of the process by which it is generated' (Kahneman and Tversky 1982B, p. 85). Strictly spoken, this hypothesis claims that subject use *given* estimates on the representativeness of the data to make probability estimations. In this sense, the second hypothesis is conceptually independent of the first. Thus, judgments *of* representativeness deal with the representativeness relation outside the context of random sampling. In contrast, judgment by representativeness refer to the judgments people make when asked to give a probability estimate.

The discussion a number of examples, as listed in Table 6, may further clarify the distinction between judgments of and by representativeness. In the remainder of this subsection I will touch upon most of the studies mentioned in the table, while leaving a (more elaborate) discussion for later sections. Let me first discuss a simple example: the study on the law of small numbers discussed in section 1.4.1. Recall that subjects were asked to estimate the probability that a significant outcome with a group of 20 subjects, would also be significant for a sample of 10 subjects. The estimates given were much higher than theory would predict. Apparently, the subjects thought that the smaller sample of 10 subjects was very much alike its 'parent population' of 20 subjects. Notice that the third type of the representativeness relation is at play here, with the group of 20 as a class and the group of 10 as a subset.

Another clear example is presented by the Tom W. study. In the Tom W. study, the 'similarity group' is explicitly asked to make a judgment of representativeness, as they are asked how much the description of Tom (instance) resembles a typical graduate student (class). Notice that this is a type 2 representativeness relation with a reversed direction. Furthermore, this judgment clearly is given outside the context of random sampling. A 'Prediction Group' was asked to estimate the probability that Tom W. is now a student in one of the designated graduation areas. Their estimates were very similar to those of the similarity group, suggesting that subjects in the latter group only judged by representativeness.

Table 6
Judgments of and by Representativeness

Study	**Judgment Of Representativeness**	**Judgment By Representativeness**	**Result**
Law of Small Numbers	Similarity (Type 3)	-	Incorrect
Bags and Chips	Similarity (Type 2)	By Similarity	Conservativeness/ Base-rate neglect
Deck and Cards	Similarity (Type 2)	By Similarity	Conservativeness
Tom W.	Similarity (Type 2), 'Similarity Group'	By Similarity, 'Prediction Group'	Base-rate neglect
Engineer-Lawyer Study	Similarity (Type 2)	By Similarity	Base-rate neglect
Medical Screening	Causal Relation (Type 4), only likelihoods	By Causality	Base-rate neglect
Suicide (incidental)	Causal Relation (Type 4), only likelihoods	By Causality	Base-rate neglect
Suicide (causal)	Causal Relation (Type 4), likelihoods and base rates	By Weighing	More or less correct

Notice that the set-up of the Tom W. study is very similar to the Engineer-Lawyer study. In the latter study, a similarity group is missing, but subjects of course still need to make an estimate of $P(D|EG)$, with *D* as the personality description and *EG* as engineer, before they can estimate $P(EG|D)$. In this respect, remark that study of judgments of representativeness checks whether estimates of $P(D|EG)$ are accurate. For example, in the case of Tom W., Tversky and Kahneman have argued that these estimates are flawed because of the notoriously low quality of the personality descriptions.

The study of judgments by representativeness checks whether estimates of $P(EG|D)$ are accurate, regardless whether the underlying estimates of $P(D|EG)$. This is most clearly seen from the studies in which the value of $P(E|H)$, with *E* for evidence and *H* for hypothesis, is already given, such as the studies dealing with the Suicide Problem. In these studies, the role of judgments of representativeness is different than in the aforementioned studies. The crucial point here is whether the base rates and conditional likelihoods are interpreted causally by the subjects. If both are interpreted in a causal manner, the judgment by representativeness that follows tends to give a more or less correct probability estimate.

In what follows I will evaluate Gigerenzer's criticism on the representativeness heuristic. Gigerenzer's treatment of the distinction between judgments of and by representativeness differs from mine. Specifically, he has labeled the four basic cases of judgments of representativeness as cases of judgment by representativeness. As a result he fails to assign the distinction between the two types of representativeness properly. In the coming sections I will argue that a number of Gigerenzer's claims against the representativeness heuristic can be invalidated because of his improper account of the heuristic.

3.2 Representativeness as a Restatement of Base-Rate Neglect

Gigerenzer has fiercely criticized the representativeness heuristic. Foremost, he has argued that the representativeness heuristic only gives a restatement of the phenomenon of base-rate neglect. This claim is made up of the following two subclaims, as found in Gigerenzer and Murray (1987, pp. 153-155):

1. Judgments of representativeness are in all cases is reducible to (conditional) likelihood;
2. Giving judgments by representativeness equals committing base-rate neglect;

Erroneously, Gigerenzer and Murray treat the four basic cases listed above as covering judgments by representativeness. As a gesture of charity, I have interpreted the first claim as addressing judgments of representativeness. I will now first evaluate these two claims and only afterwards I will discuss the implications that Gigerenzer has attached to these claims. I will start off with the first claim.

3.2.1 Representativeness in all cases is reducible to (conditional) likelihood

Gigerenzer and Murray (1987, p. 155) argue that in all the four basic cases of judgments of representativeness, representativeness is reducible to (conditional) likelihood. Let me clarify this claim by discussing the case of Jack, whose personality description is deemed to be very stereotypical of an engineer. The representativeness relation here comes down to the second basic class of judgments of representativeness, with engineers as a class and Jack as a sample of that class. Kahneman and Tversky (1982B) claim about this second class that 'an instance is representative of a category if it has the essential features that are shared by members of that category and does not have many distinctive features that are not shared by category members' (p. 86).

Gigerenzer and Murray (1987, p. 154) now remark that the directional relation from class to instance in representativeness relations is the same as the relation from a population or hypothesis to a single event for likelihood relations. Consequently they wonder whether representativeness is not the same as assuming some subjective multidimensional distribution *H*, for which $P(D|H)$ is the likelihood of a single event *D*. For example, Figure 4 shows a *random* probability distribution of engineer properties.[23] As the personality description of Jack shows many stereotypical engineer properties, Jack is located at the far end of the probability distribution, which is nothing different from saying that $P(D|EG)$[24] is fairly high. A similar graph can be drawn for $P(D|L)$[25], with lawyer instead of engineer properties on the horizontal axis, such that $P(D|L)$ is located at the near end of the graph.

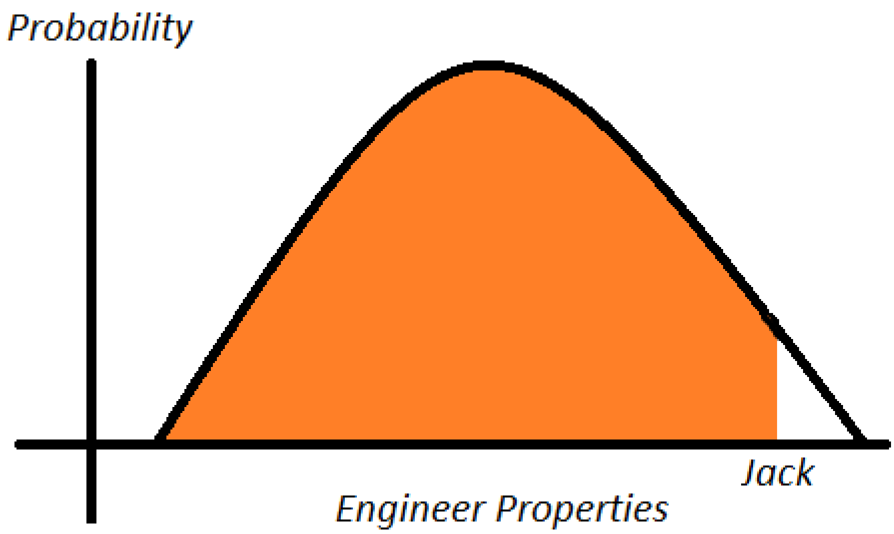

Figure 4

As I have shown by discussing the case Jack, judgments of representativeness concerning the second basic case can be described in terms of conditional likelihood. For the other basic cases, similar descriptions can be given. As Gigerenzer and Murray (1987, p. 155) remark, the fourth case is only a special case of the first. And similar to the second case, the first case is only a more specific version of the third case. It appears that Gigerenzer and Murray (1987) are right in claiming that 'representativeness in all cases is reducible to [conditional] likelihood' (p. 155), given that representativeness refers to judgments of representativeness here.

3.2.2 Representativeness as a redescription of the phenomenon of base-rate neglect

In the last subsection, I have argued that judgments *of* representativeness can be described in term of conditional likelihood. I will now continue by discussing Gigerenzer and Murray's second claim, which covers judgments *by* representativeness. This claim implies that all cases of base rate neglect can be interpreted as caused by judgments by representativeness and vice versa that all cases of judgments by representativeness lead to base rate neglect. Gigerenzer (1991A) has argued that this would make the representativeness heuristic a 'largely undefined concept [that] can post hoc be used to explain almost everything' (p. 102).

[23] To be interpreted as having engineer properties without having distinctive features that are not shared by other engineers.
[24] The conditional probability that Jack is an engineer *EG* given the description *D*
[25] The conditional probability that Jack is a lawyer *L* given the description *D*

Let me start off with discussing the claim that all cases of base rate neglect can be interpreted as caused by judgments by representativeness. Tversky and Kahneman would probably oppose to the identity of judgments by representativeness and the phenomenon of base-rate neglect. In 'Evidential Impact of Base Rates', Tversky and Kahneman (1982B) argue the following:

Predictions by representativeness or similarity are generally insensitive to base-rate frequencies. However, the phenomenon of base-rate neglect is far more general, since it also occurs in judgments that cannot be readily interpreted in terms of representativeness (Hammerton, 1973). (*p. 154)*

Recall that in the study by Lyon and Slovic (1976), results to Hammerton's studies on the 'Disease Problem' and the 'Engine Crack Problem' all show base-rate neglect. In these two studies subjects were asked to estimate the probabilities that an engine is broken or that a person carries a disease, given a positive result. Subjects are told that $P(H)$, the probability of having the disease or engine problems is 0.01, such that $P(\neg H)$is 0.99. Moreover, $P(P|H)$, the probability of getting a positive result *P* when in fact having the disease/engine crack is .9, while $P(P|\neg H)$ is .01.

Normally, one would expect that *H* causes *P* and ⌐H not to cause *P*, such that *P* is a 'diagnostic' piece of evidence with respect to *H* (and ⌐*H*) as Tversky and Kahneman would call it (1977B, p. 1-1). In other words as a judgment of representativeness, we have a version of the fourth basic case, with *H* (⌐H) as the system and *P* (not) following from it as a consequence. To both problems, the median estimates were .90, which is the same as the value for the accuracy of the test in case the person carries the disease or the engine is broken. Apparently, $P(H|P)$ is calculated solely by use of $P(P|H)$, similar to judgments by representativeness. Indeed, one can speak of a judgment by representativeness here, as *P* 'reflects the salient features of the process by which it is generated' (Tversky and Kahneman 1982B, p. 85). It appears that although Hammerton's problems are not 'readily interpreted in terms of representativeness' according to Tversky and Kahneman (1982B, p. 154), they can certainly be interpreted in that way. The same holds for another candidate, the Suicide Problem, which I will discuss below.

I shall continue the discussion by assessing the claim that the use of judgments by representativeness always leads to base rate neglect. If this were indeed true, it would hold that judgments by representativeness entail that that $P(H|E)$ is equated with $P(E|H)$, such that base rates are neglected. Indeed, in a lot of studies discussed in this thesis, judgments by representativeness and base-rate neglect coincide. The Taxicab and Suicide Problem are noteworthy exceptions though.

In the first version of the Suicide Problem presented in section 2.4.5, the median estimate was 0.75. This equals $\frac{P(D|S)}{P(D|S)+P(D|M)} = \frac{0.15}{(0.15+0.05)}$, with $P(D|S)$ and $P(D|M)$ the probability of death by suicide *D* for singles *S* and married persons *M* respectively. This estimate shows entire neglect of base rates.

The estimates to the Suicide Problem can be explained in terms of judgments by representativeness. For whatever reason, people may believe that being single causes higher suicide rates than being married (relative to other non-natural causes of death). In other words, death by suicide reflects the salient features of the process by which it is generated to a higher degree for singles than for married persons.

What I have shown is that the results to the first version of the Suicide Problem can be interpreted as resulting from a judgment by representativeness. As one may recall, the results to the second version of the Suicide Problem hardly show any base-rate neglect at all. In the second version of the problem subjects are told that 80% of suicide attempts are made by girls, and 20% by boys. Thus, girls (boys) function as a causal system with a relatively high (low) number of suicide attempts as a consequence. These judgments of representativeness are already reduced to conditional likelihoods, as we are told that $P(G|S)$, the probability of a girl *G* given a suicide attempt *S* is 80%.

$P(G|S)$ and $P(B|S)$, the probability of a boy B given a suicide attempt, are not the only conditional likelihoods that are given in this problem though! Subjects are also told that 'the percentage of suicide attempts that result in death is 15% among boys and 5% among girls' (Tversky and Kahneman 1977a, p. 178). Thus, boys (girls) function as causal systems having a high (low) rate of succeeded suicide attempts as a consequence. As conditional likelihoods we have $P(D|B)$ and $P(D|G)$, the probability of death given a suicide attempt of a boy or a girl respectively.

Thus, instead of one, two judgments *of* representativeness can be identified in the second version of the Suicide Problem. Now, assume that a subject will judge by representativeness, in the sense that she estimates the probability of an uncertain event to the degree that it reflects the salient features of the process by which it is generated. Recall that probability of a girl, given the knowledge that someone died through suicide $P(G|D)$ is asked for. Bayes' theorem gives:

$$P(G|D) = \frac{P(G|S) * P(D|G)}{P(G|S) * P(D|G) + P(B|S) * P(D|B)} = \frac{0.8 * 0.05}{0.8 * 0.05 + 0.2 * 0.15} = \frac{3}{7} \text{ [Eq. 6]}$$

As has been shown, the process or causal system girl generates two salient features here. It generates a higher number of suicide attempts, but at the same time also a lower number of succeeded attempts relative to boys. How is the subject going to use this information in a judgment (by representativeness)? Apparently, as the median result of .5 shows, subjects succeed in weighing the pieces of information, such that neither the base rates ($P(G|S)$ and $P(B|S)$), nor the conditional likelihoods $P(D|G)$ and $P(D|B)$are (grossly) neglected.

Thus, from the second version of the Suicide Problem it can be very well defended that judgments by representativeness do not necessarily lead to base rate neglect. Notice that Gigerenzer and Murray (1987) argued that the use of a representativeness heuristic for probability revision simply means that subjects use the likelihood in Bayes' theorem but not the prior probabilities. What Gigerenzer and Murray overlooked is that subjects do focus on the prior probabilities when these can be interpreted as conditional likelihoods. More generally spoken, Gigerenzer and Murray's error to treat the four basic cases of judgments *of* representativeness as such, now backfires on them. Because of this mistake, they cannot account for the possibility that multiple judgments of representativeness can underlie a judgment by representativeness.

From this subsection, one can conclude that although it is in principle possible to redescribe base-rate neglect in terms of judgments by representativeness, the reverse is not possible in all cases. That is, not all judgments by representativeness lead to base-rate neglect as the Suicide Problem shows. Therefore, Gigerenzer and Murray's second claim that 'representativeness is a redescription of the phenomenon of base-rate neglect' is incorrect (Gigerenzer and Murray 1987, p. 155). Rather, claim 2 should be replaced by the weaker claim 2': 'The phenomenon of base-rate neglect can be described in terms of representativeness'.

3.2.3 The Implications of Gigerenzer and Murray's claims

In the preceding two subsections I have discussed two claims of Gigerenzer and Murray (1987) covering the representativeness heuristic. Against the first claim that representativeness is reducible to (conditional) likelihood I have found no objections. Towards the second claim I have posed some reservations, arguing that it should be changed. In this subsection I will introduce the implications that

Gigerenzer and Murray have attached to the original claims. In the following subsections I will evaluate these implications, in which I will also argue how my proposed change to Gigerenzer and Murray's second claim affect the validity of these implications.

The first implication is rather simple. If representativeness is merely a restatement, it cannot be explanatory. As Gigerenzer (1996, p. 594) comments, arguing that opium makes you sleepy because of its dormative properties does not provide an explanation. Likewise, claiming that people commit base rate neglect because of the representativeness heuristic offers no genuine explanation. In section 3.3, I will discuss this topic in more detail. Drawing on the urn problems introduced in section 1.4.1, I will show that the representativeness heuristic can be explanatory.

The second implication is that the representativeness heuristic is largely an undefined concept. As a result, all cases of base-rate neglect can be reduced to the representativeness heuristic. For this reason, Gigerenzer (1996) argues that the theory on the representativeness heuristic does not give falsifiable predictions. Moreover, no underlying cognitive process labeled representativeness is specified, despite the stated goal to study cognitive processes. Consequently, no theory is provided concerning the relation of this process 'to any specific content, context, or representation of numerical information' (Gigerenzer 1996, p. 594). In section 3.4 I will discuss this implication in more detail. I will extensively discuss Tversky and Kahneman's defense, in which they argue that Gigerenzer's criticism misses the point, as representativeness is assessed empirically.

The third implication, as pointed out by Gigerenzer and Murray (1987, p. 155), is that representativeness stems from the Bayesian framework, such that it has inherited the attributes of Bayesian thinking. These attributes are the following. First, Bayesian thinking ignores the role of the passions and only focuses on cognitive aspects. Second, it neglects the process of search for information. Third and fourth, Bayesian thinking is insensitive to the content and context of a problem. In section 3.5 I will discuss these attributes and argue that although the representativeness heuristic can be associated with the first attribute, it cannot be so with the others. One can in fact wonder whether these attributes are part of Bayesian thinking in the first place.

3.3 Explanatoriness of the Representativeness Heuristic

As introduced in the previous subsection, Gigerenzer has criticized the representativeness heuristic for not being explanatory. In this subsection I will show that the representativeness heuristic can be explanatory by drawing on the urn problems discussed in section 1.4.1.

Recall that the estimates to the deck-and-cards problem were conservative relative to the Bayesian estimate. These results seem to be in sharp contrast with the base-rate neglect studies conducted within the heuristics and biases program. Or as Gigerenzer and Murray (1987) have commented:

It is ironic that the whole phenomenon of conservatism disappeared when in the early 1970s Daniel Kahneman and Amos Tversky posed Bayesian problems with a content different from bookbags and poker chips. Subjects no longer seemed to reason conservatively about the new problems; indeed, they even seemed to neglect the prior probabilities. (p.150)

Gigerenzer et al. (1989,) also comment on this issue:
The question of why people seemed to be conservative during the 1960s and anti-conservative after 1970 has not yet been answered, if only because it was almost never posed. It should be very disturbing that established facts suddenly do an about face. But the new facts were instead enthusiastically

received as revelatory of underlying mental heuristics, and the opposite facts largely ignored as too old to be true. (p. 219)

Gigerenzer and Murray (1987) construct a dichotomy between experiments showing base rate neglect and those showing conservatism. This dichotomy is unfounded though, as it can be shown that both results are driven by judgments by representativeness. In other words, the representativeness heuristic can both explain conservativeness in the early urn problems as well as base rate neglect in the later base-rate neglect studies. To see this though, one needs the conceptual distinction between judgments of and by representativeness that Tversky and Kahneman (1982B) only introduced ten years after the publication of the results to their deck-and-cards experiment (Kahneman and Tversky 1972B).

Recall that in the deck-and-cards problem, a number of cards either marked 'X' or 'O' is randomly drawn from either one deck of cards. Deck *A* predominantly has cards marked 'X', while deck *B* mostly contains of cards marked 'O'. Being informed that a random draw gives a specific sample ratio (with predominantly cards marked 'X'), subjects have to give an estimate of the probability that the sample is drawn from deck *A*.

Subjects will now typically make a judgment of representativeness to calculate this probability. This implies that subjects expect samples to be highly similar to their parent population. Indeed, because of the resemblance of the draw 5:1 and the population proportion of deck *A*, subjects are much more confident that this sample is drawn from deck *A* than for other more probable samples. In other words, subjects assign a relatively high value to $P(D|A)$, the conditional probability of drawing 5:1 (*D*) given deck *A*.

The judgment of representativeness is followed by a judgment *by* representativeness, which entail that subjects will use (their estimation of) $P(D|A)$ and $P(D|B)$ to calculate $P(A|D)$. Notice that in the formulation of this problem base-rate neglect is justified as the base rates for both decks are identical, such that they cancel out in Bayes' theorem. Base-rate neglect is however likely to occur in this problem even if the base rates would differ between two types of decks. Indeed, in the urn problem of Phillips and Edwards (1966, p. 348), variation of prior probabilities[26] did not affect the subjects' estimates, which showed 'conservatism'.

Thus, both the conservative estimates in the urn problems as well as the neglect of prior probabilities in Tversky and Kahneman's later studies can be accounted for with the representativeness heuristic. The difference between the two types of studies lies in the underlying judgment of representativeness. In the case of the urn problems, these judgments are faulty and therefore lead to final estimates which are lower than the Bayesian estimate. In case of Tversky and Kahneman's base-rate neglect studies, these estimates may or may not be faulty, but lead to an estimate that is higher than the Bayesian estimate due to the neglect of base rates.

Contrast to what Gigerenzer has argued, no gulf exists between the 1960's and the 1970's studies on the use of Bayes' theorem. In arguing towards this conclusion, I have emphasized the explanatory role that the representativeness heuristic has towards both conservativeness and base-rate neglect. Thus, contrary to Gigerenzer, I would argue that the representativeness heuristic is explanatory, explaining the link between the two phenomena, which was according to Gigerenzer himself not addressed before. Because Gigerenzer has not carefully followed the distinction between judgments of and judgments by representativeness in his writings, he has not been able to expose the roles these two types of judgments have in the urn problems discussed in this subsection.

---

[26] That is, variation in the number of bags with predominantly red or blue chips

3.4 Limits to the Application of the Representativeness Heuristic

Gigerenzer (1996) has argued that representativeness is an undefined concept. A corollary to this is that basically any case of base-rate neglect can be described in terms of representativeness. This leaves to wonder how the scope of the representativeness heuristic is limited. Or as Gigerenzer (1996) would put it: how does a theory on representativeness produce falsifiable predictions? Another worry that Gigerenzer (1996) as attached to the vagueness of the representativeness heuristic, is that the underlying cognitive process is left unmentioned, including the relation of this process to variables such as the context and content of a problem. In this subsection I will address these concerns in more detail. I start off with discussing the scope of the representativeness heuristic.

3.4.1. Representativeness Defined Empirically

Tversky and Kahneman (1982B, p. 154) have argued that the phenomenon of base-rate neglect is far more general than the representativeness heuristic. As I have failed to show the contrary, I cannot refute the claim that base-rate neglect can always be reduced to judgments by representativeness. But how then is its applicability limited?

In their 1996 reply to Gigerenzer's repeated criticism towards the heuristics and biases approach, Tversky and Kahneman explicitly dealt with Gigerenzer and Murray's criticism on the representativeness heuristic. Importantly, Kahneman and Tversky (1996) do not deny that representativeness is a largely undefined concept as Gigerenzer understands it to be. Instead, they argue that this criticism misses the point as 'representativeness can be assessed empirically; hence it need not be defined a priori' (Kahneman and Tversky 1996, p. 585). As an example of such empirical assessment, Kahneman and Tversky (1996, p. 585) introduce the so called outcome-ranking paradigm. The clearest example within this paradigm is the Tom W. study. Recall that three different groups were asked to rank a set of fields of study by base rates, representativeness or probability (Table 4).

The similarity group in the Tom W. study was explicitly asked to make a judgment of representativeness. The second case judgments of representativeness applies here, with *M* replaced for a field of graduate specialization and *X* replaced for Tom W.('s personality description). Notice that these judgments of representativeness by this group of subjects are not reflecting biases, as they were not told that the personality description is fallacious.[27] The prediction group did make errors though. As one can see from Table 4, the rankings of outcomes by representativeness are nearly the same as the ranking of outcomes by probability, reflecting that subjects both overlooked the base rates and the bad quality of the provided information. Apparently, subjects used their judgments of representativeness to estimate probabilities, which is what judgments by representativeness entail.

Kahneman and Tversky (1996, p.585) argue that no theoretical model is needed to test the hypothesis whether judgments under uncertainty are mediated by representativeness. The only requirement is that representativeness is used as a tool for probability judgments and not the other way around. Furthermore, representativeness is defined empirically, as done by the similarity group in the Tom W. study. Subsequently, these estimates are compared to the probability judgments of another group of people, the prediction group in the Tom W. study. Tversky and Kahneman emphasize that it is the understanding of the subject of similarity and probability that determine these rankings, just like a subject in perceptual research decides what meaning to attach to a concept such as loudness. This is all in line with the earlier discussed claim that Tversky and Kahneman's research uses people's intuitions to test whether people's intuitions are valid. Subjects' intuitions are drawn upon to give

[27] If they would have been told so, one would have expected that the 'similarity rank' would be more reflective of the base rates (second column, Table 4)

judgments of representativeness, to give probability estimates and to decide whether a probability estimate is normatively valid or not.

Tversky and Kahneman argue that representativeness does not have to be defined a priori, as it is defined operationally by the subjects' ranking. This procedure applies especially to the outcome-ranking paradigm, but is not limited to it. In 'Judgments of and by Representativeness', Kahneman and Tversky (1982B, p. 88) argue that the following assumptions need to be satisfied in order to test whether judgments of probability are mediated by assessments of and by representativeness:

1. *The relation "X is (very, …, not at all) representative of M" can be meaningfully assessed by judges.*
2. *These assessments should not be based on impressions of probability or frequency, which are to be explained by representativeness.*
3. *The relation of representativeness has a logic of its own, which departs systematically from the logic of probability*

The first condition concerns judgment of representativeness, such as made by the similarity group in the Tom W. study. The second condition simply states that in judgments by representativeness $P(E|H)$, with *E* for evidence and *H* for hypothesis is used to estimate $P(H|E)$and not vice versa. The third condition states that deviations from the logic of probability due to (judgments of and by) representativeness should be systematic. Thus, as explained by Kahneman and Tversky (1996), representativeness is defined empirically, not a priori, by the subjects in the experiments. It is merely the inclination of the subject to interpret a relation between an instance *X* and a class *M* as representativeness which is the decisive factor for determining whether a judgment is being led by representativeness or not. This inclination is of course influenced by factors such as the perception of the content and context of a problem by the subject. I will return to this issue in sections 3.5.4 and 3.5.5.

In some of the problems, such as Tom W. or the Engineer-Lawyer-study, a relation of representativeness is easier to point out than in other problems such as Hammerton's. This observation clears up why Tversky and Kahneman argued that base-rate neglect occurs in judgments that cannot be *readily* interpreted in terms of representativeness. It is not the experimenter, but the subjects who apparently not readily see a representativeness relation in Hammerton's problems. Representativeness is ultimately measured through subject's opinions. Whether relations of representativeness apply to a certain problem is therefore eventually decided upon by the subjects in an experiment. In that sense, the scope of the representativeness heuristic is limited.

Contrary to what Gigerenzer has argued, Tversky and Kahneman's theory of representativeness does make falsifiable predictions. Once a relation is judged to be strongly representative, this relation should have a logic of its own, systematically departing from the logic of probability as Kahneman and Tversky (1982B, p. 88) have demanded. Originally, the representativeness heuristic entailed that 'probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B' (Tversky and Kahneman 1974, p. 1124).

In the 1974 formulation of the representativeness heuristic, a judgment by representativeness with the neglect of base rates as a consequence would follow whenever a relation of representativeness is found. With the results to the Taxicab and the Suicide problem though, which also contain relations of representativeness, Tversky and Kahneman have falsified their earlier account of the representativenss heuristic. Apparently, whenever both the base rates and the conditional likelihoods are interpreted causally, base rates are not neglected. As a consequence, Tversky and Kahneman (1982B) changed the views on representativeness as voiced in for example Tversky and Kahneman (1974) accordingly. This change to the views on the representativeness heuristics shows that it does give falsifiable predictions: namely, whenever judgments that follow from a judgment of representativeness do not systematically depart from the logic of probability, the theory on the representativeness heuristic is falsified.

3.4.2. The coexistence of the representativeness heuristic and mental models

In this subsection I would like to return to the case of Dick, as introduced in section 2.3.1. Gigerenzer (1991A, p. 100) has used the case of Dick to argue that the representativeness is inferior to an explanation appealing to mental models. In arguing so, he has created a dichotomy between explanations bearing on the representativeness heuristic and those that use mental models. Here I will argue that the creation of such a dichotomy is unfounded, as Tversky and Kahneman have actively resorted to both mental models and representativeness in their explanations.

Recall that in the Engineer-Lawyer experiment, the description of Dick was constructed to be uninformative. Thus, probability estimates should have been based on base rates only. Nevertheless, Kahneman and Tversky (1973) found almost complete neglect of base-rates for the case of Dick. In their rerun of the experiment, Gigerenzer, Hell and Blank (1988) found considerable neglect of base rates for the uninformative description of Dick, for both verbal and visual random sampling. Gigerenzer, Hell and Blank (1988) point out though that a comparison with other studies shows that Bayesian performance for estimates to Dick's description are weaker the more descriptions are shown to the subject. Especially when subjects are only shown one description, subjects' estimates for Dick do not deviate from the Bayesian estimates. This brings Gigerenzer, Hell and Blank (1988) to the conclusion that the order of presentation is a critical factor for the evaluation of non-diagnostic information.

Gigerenzer (1991A) interprets the importance of order in the probability estimates for Dick as an argument against the representativeness heuristic. He simply states that 'this result should not occur if the intuitive statistician operates with a representativeness heuristic' (Gigerenzer 1991A, p.100). Instead, Gigerenzer (1991A) argues, a mental model (dealing with profession guessing) is activated when subjects encounter the first personality description. This mental model looks for cues, such as hobbies and political attitudes, which make it more probable that a person has a certain occupation. Subsequently, upon encountering the description of Dick this mental model is used, which apparently does not take base rates into account. This mental model acts like a shortcut procedure, as introduced in section 2.3.2. When Dick is encountered first though, the mental model is not activated and base rates are taken into account.

Gigerenzer (1991A) does not explain why mental models and the representativeness heuristic cannot coexist. Furthermore, there is no reason to think they cannot when one draws upon the work of Tversky and Kahneman, who have argued that 'the properties that make formally equivalent problems easy or hard to solve appear to be related to the mental models, or schemas, that the problems evoke' (Kahneman and Tversky 1982A, p. 130). Apparently, the Engineer-Lawyer study activates a mental model in which the personality descriptions first encountered act as an anchor for later descriptions. One can conclude from this subsection that although Tversky and Kahneman have not specified mental models in the context of the representativeness heuristic, they have never claimed that the two are incompatible, unlike what Gigerenzer (1991A) has argued. Why Tversky and Kahneman have not specified mental or process models will be the concern of next subsection.

3.4.3. Process Models

Gigerenzer (1996) is not impressed by Kahneman and Tversky's account of the representativeness heuristic. In his reply to Kahneman and Tversky 1996, Gigerenzer (1996) argues that 'progress can be made only when we can design precise models that predict when base rates are used, when not, and why' (p. 595). The representativeness heuristic does not give such a precise model, as one cannot know beforehand whether subjects will observe a relation of representativeness. And even if they might not do so 'readily', they might still neglect base rates as in Hammerton's problems. Gigerenzer concludes that there is a need for models that make surprising and falsifiable predictions, which also expose the mental processes that account for valid and invalid judgments.

Kahneman and Tversky (1996) have not denied that representativeness is an indeterminate concept, as Gigerenzer (1996, p.596) claims it to be. Instead in the postscript to their 1996 article (Kahneman and Tversky 1996, p. 591), which replies to Gigerenzer (1996), they argue that the conditions that Gigerenzer put up are too strict. A lot of good psychology, such as the Gestalt rules for good continuation and similarity, do not satisfy Gigernzer's conditions. Tversky and Kahneman do not think that the construction of process models form the right strategy to forward the study of judgmental heuristics. Again they refer to Gestalt psychology, where models soon got outdated, while its qualitative principles still stood firm. For similar reasons, Tversky and Kahneman argue that more progress will be made when building upon concepts such as representativeness.

Clearly, Tversky and Kahneman, who have Gigerenzer on their side with the goal to 'understand the cognitive processes that produce both valid and invalid judgments' (Kahneman and Tversky 1996:582), do not think that formulating precise models is the way to progress in studying judgment. Although they do not oppose modeling, they think that process models are likely to be premature and therefore prone to become outdated very soon. In his postscript, Gigerenzer (1996, p. 596) replies that the difference in research strategy is the dividing issue between him and Tversky and Kahneman.

The positions of Gigerenzer and Tversky and Kahneman can both be defended. As Gigerenzer (2010, p. 733) claims, other scientific disciplines, such as biology, economics, or physics are much more concerned with theory construction and integration than psychology. As a result, the relation between the various concepts in the dozens of theories that psychology has come up with, is quite unclear. This does not only hinder progress in the science of psychology itself, but also leads to a reduced interest for psychological theories in neighboring disciplines such as economics.

Tversky and Kahneman, on the other hand, have good reasons not to engage in the construction of more precise theories. Firstly, with the fast accumulation of new data in this relatively young scientific enterprise, theories get outdated quickly. Secondly, experimental evidence for the process models of cognition put forward by Gigerenzer and associates is still lacking, such that these models remain speculative (Fiedler and von Sydow 2015, p. 152-153). Thirdly, the use of precise models might cast serious disagreement among scientists, thereby reducing the attractiveness of the research program. As has been widely acclaimed, even by Gigerenzer (1991A, p. 85), the heuristics and biases program had an enormous impact in psychology as well as in other disciplines. It is very much doubtful whether the research program could have got so popular, if it had started with advocating theories that many of its potential supporters would feel uncomfortable with.

In this subsection I have explained that the scope of the representativeness heuristic is limited empirically and not theoretically, as Gigerenzer appears to require. Contra Gigerenzer, I have argued that the theory on the representativeness heuristic does give falsifiable predictions. Moreover, I have shown that Tversky and Kahneman do not object to Gigerenzer's criticism that the representativeness heuristic is mostly undefined. Tversky and Kahneman have rather stressed that they do not share Gigerenzer’s desiderata for a theory in psychology. This has been a reason for Gigerenzer to relabel the dispute between him and Tversky and Kahneman in terms of different research strategies. I have argued that it is mostly an empirical matter whether one strategy is stronger than the other. There are certainly no knock-down theoretical arguments to favor one approach over the other.

3.5 The Attributes of Bayesian Thinking

Building on the claim by Gigerenzer and Murray (1987, p. 155) that representativeness is reducible to (conditional) likelihood, one can come to argue that representativeness stems from the same framework as Bayes’ theorem. With this claim of Bayesian origination, Gigerenzer and Murray mean that psychological explanations come from the vocabulary of probability theory and Bayes’ theorem particularly (Gigerenzer and Murray 1987, p. 150). This claim is valid insofar that if judgments of representativeness are indeed equal to an estimation of conditional likelihood, one of the constituents of Bayes’ theorem, then it follows that representativeness is at least compatible with the same Bayesian framework as Bayes’ theorem. Notice in this respect that I drew on Bayes’ theorem in explaining that judgments of representativeness do not lead to base rate neglect when prior probabilities are presented as conditional likelihoods in section 3.2.2. In conclusion, I see no reason to refute the claim of Bayesian origination.

3.5.1. The Tools to Theory Heuristic

I will now argue how Gigerenzer and Murray (1987) substantiate the claim of Bayesian origination and the subsequent ‘claim of inheritance’. This claim says that the representativeness heuristic has inherited the major attributes of Bayesian thinking, which will be introduced below. Gigerenzer and Murray draw upon what Gigerenzer and Sturm (2007) would later come to call the ‘Tools to Theory Heuristic’ to make their point. I will not criticize this heuristic by itself here. My only purpose with introducing Tools to Theory Heuristic here is to explain how Gigerenzer and Murray got to their position. What will be subject of evaluation in this subsection is the just mentioned claim of inheritance.

The claim of Bayesian origination of representativeness is an illustration to the view of Gigerenzer and Murray (1987) that theories in psychology about human cognition often follow from the (statistical) tools used by psychologists. For instance, progress in computation gave rise to theories that modeled the mind as a computational system, which got especially popular in the cognitive science from the 1960s on (Rescorla 2015). Gigerenzer’s point is that the adaptation from tools to theories often proceeds unconsidered.

Gigerenzer ‘Tools to Theory Heuristic’ is best compared to the better known phenomenon of Theory Ladenness of the Observation. In the latter, theories determine what data are sought after and what tools are used for that purpose, but they also determine how the data are to be interpreted. Amongst other effects, this may imply a confirmation bias, as the data are not theory-neutral but bound to be constructed along the lines of the theory.

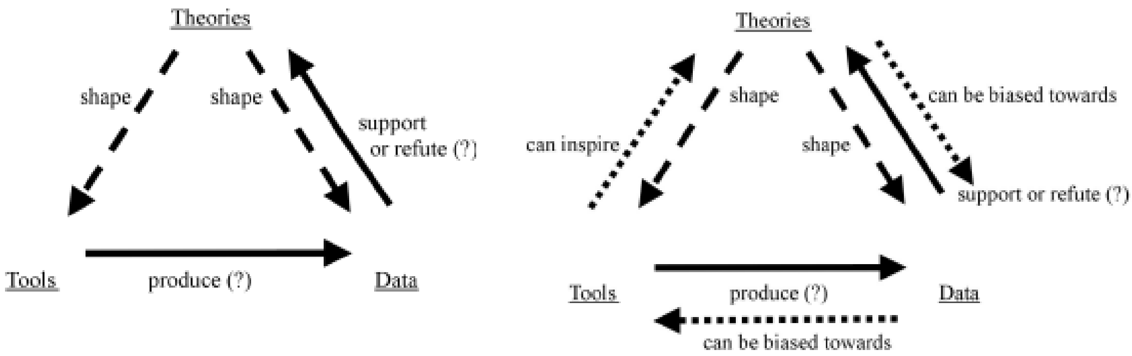

Figure 5A (left) 'Theory Ladennes of the Observation' and 5B (right) 'Tools to Theory Heuristic'

The Tools to Theory Heuristic adds to Theory Ladenness of the Observation that theories are often inspired by the tools favored by scientists and that only data that fit the favored tools are used. Thus, Theory Ladenness of the Observation and Tools to Theory Heuristic together turn confirmation bias into something circular: tools, theories and data are selected in such a way that they tend to justify one another in a circular way. Figure 5A and 5B, taken from Gigerenzer and Sturm (2007, p. 314) provide illustrations to both Theory Ladenness of the Observation and Tools to Theory Heuristic.

The early research in Bayesian reasoning by Edwards and other fits nicely with the Tools to Theory Heuristic. This research would not have been possible without the development of theories and tools in subjective probability. Edwards concluded that people reason in accordance with Bayes' theorem, the statistical tool that inspired his research. Tversky and Kahneman rejected this conclusion, arguing that man is not a Bayesian at all. Instead, judgments are made on the basis of the representativeness heuristic. Nevertheless, the Tools to Theory Heuristic can in principle apply here as well, because representativeness stems from Bayes' theorem as the claim of Bayesian origination describes. The Tools to Theory Heuristic entails that data and theory are biased towards the tools used. With the Tools to Theory Heuristic, Gigerenzer and Murray (1987) attempt to substantiate the claim of Bayesian origination and the claim of inheritance.

These major attributes that the representativeness heuristic has allegedly inherited from the Bayesian framework are the following:

a) It emphasizes rationality rather than passion (**Rationality Attribute)**
b) It has nothing to say about the process of information search (**Search Neglect Attribute)**
c) It operates independently of the content of a problem (**Content Insensitivity Attribute**)
d) It operates independently of the context of a problem (**Context Insensitivity Attribute**)

In what follows, I will discuss whether the representativeness heuristic really has these attributes. There are two ways to go about this. First, one might check whether the attribute is part of Bayesian thinking in the first place. Second, one might check whether the attributes really is 'inherited' by the representativeness heuristic. I will mainly draw on the second strategy, though occasionally I will rely on the first strategy. I will argue that there is sufficient reason to assume that the representativeness heuristic has the rationality attribute. Of the other attributes I will argue that they are not part of the representativeness heuristic.

3.5.2. The Rationality Attribute

There is sufficient reason to argue that Tversky and Kahneman would actually agree with the criticism put forward in the ascription of the Rationality Attribute to the representativeness heuristic. In their discussion of the affect heuristic, which describes that human judgment is led by the feelings that some piece of information triggers, Kahneman and Frederick (2002, p. 56) admit that 'the failure to identify this heuristic earlier reflects the narrowly cognitive focus that characterized psychology for some decades'. This does not seem very different from Gigerenzer's argument that the cognitive revolution in American Psychology in the 1960's had the mind pictured as an 'intuitive statistician' or a 'computer program' (Gigerenzer 1991B, p. 254 & 263). Gigerenzer only adds to Kahneman and Frederick that Tools to Theory Heuristic is at work here: the cognitive focus is a result of the changes in tools used by scientists or psychologists more specifically.

3.5.3. The Search Neglect Attribute

Let me now introduce the Search Neglect Attribute, which concerns a lack of interest in information search of the subjects in Tversky and Kahneman's studies. This attribute is based on the following claims:

b1 Representativeness stands in the framework of statistical problems (claim of inheritance)
b2 In statistical problems, thinking means mental work on prepackaged information
b3 Therefore, representativeness does not deal with the process of information search

Claim b1 has been defended already. Claim b2 may need some explanation. To ensure that a statistical problem has a unique solution, or at least to ensure that the isomorphism conditions as discussed in section 2.1 are fulfilled, much information must be explicitly given that in everyday circumstances would not be readily available. As an example, Gigerenzer and Murray (1987, pp. 164-165) refer to the Taxicab Problem. In Tversky and Kahneman's versions of the problem, base-rates are given by the experimenters. Recall that the original formulation of the problem claimed that '*85% of the cabs in the city are Green and 15% are Blue'*, while another version declared that '*85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs'* (Tversky and Kahneman 1982B). In both versions of the problem, Tversky and Kahneman have argued that these are the relevant base rates to be used as prior probabilities in Bayes' theorem. In a real-life situation though, a judge would have to decide for herself which base rates are relevant. Given that one knows that a cab was involved in a hit-and-run accident at a certain spot in town, different variables are of interest to be included in the base rates, in order to ensure that the prior probability is sufficiently fine-grained. Importantly, no formal or single rule exists for information search and judgment of relevance, such that a judge herself has to decide on which rules to apply.

For Gigerenzer and Murray (1987), the representativeness heuristic is tied to problems that allow for a full-fledged application of Bayes' theorem or some other statistical rule in which all information deemed relevant is already given. Therefore, claim b3 follows that Tversky and Kahneman's problems circumvent the process of information search as 'most if not all of the process of information search has already been done by the experimenter' (Gigerenzer and Murray 1987, p. 157).

There is ample reason to criticize the line of reasoning by Gigerenzer and Murray (1987) though and especially claim b2 seems to be exaggerated. Indeed in the Deck-and-Cards problem, the Taxicab Problem and the Suicide Problem, almost all information comes prepackaged. Matters are different though for the Tom W. problem, in which information on base rates and conditional likelihoods are not explicitly given, but are to be decided upon by the subjects. Although I have shown that there is reason to doubt whether the isomorphism condition on posterior probability is fulfilled in this problem, it is not a made out case that it is not. There are simply no clear-cut rules on the extent of guidance the subjects may receive. As a result, one may argue that there is insufficient reason to discard the results of the Tom W. study. Consequently, claim b2 should be weakened to b'2 'in statistical problems, thinking *involves but is not restricted to* mental work on prepackaged information'. There is simply no reason to think that the representativeness heuristic only works when all information is 'prepackaged'. Obviously, claim b3 does not follow from b'2.

Claim b'2 is also more faithful to Tversky and Kahneman's own writings on information search. According to Kahneman and Tversky (1982B, p. 88) the complex process of the evaluation of an uncertain event or the prediction of an unknown quantity comprises three stages, namely interpretation of the problem, a search for relevant information, and the choice of an appropriate response. This complex process can be compared with 'the operation of a flexible computer program that incorporates a variety of potentially useful subroutines' (Kahneman and Tversky 1982B, p. 88). Tversky and Kahneman continue that the representativeness heuristic is one of the procedures that can be drawn upon to retrieve, interpret, and evaluate information. Thus, unlike Gigerenzer and Murray's claim, the representativeness heuristic does have a role in the process of information search. For example, in the Tom W. problem, the subjects in the prediction group do not put much effort in retrieving the appropriate base rates, as the similarity between Tom W. personality description and some of the graduate specialization areas are so outspoken.

3.5.4. The Content Insensitivity Attribute

The Content Insensitivity Attribute entails that the representativeness heuristic, pinned on the structure of Bayes' theorem, is insensitive to content. Kahneman and Tversky (1996) have classified this claim together with the context insensitivity attribute as the 'the most serious misrepresentation of our position' (p. 583). Let me try to explain how Gigerenzer and Murray (1987) substantiate their claim with respect to the neglect of content. The argument that he provides for the insensitivity to content goes as follows:

- c1 Problems that seem to have a Bayesian structure should be solved using Bayes' theorem, independent of the actual content
- c2 Representativeness stems from the same framework as Bayes' theorem (claim of inheritance)
- c3 Therefore, people use a representativeness heuristic for all contents in the evaluation of uncertainty

Essentially, Gigerenzer and Murray (1987, p. 155-156) argue that the representativeness heuristic has inherited the content-blindness of its Bayesian parent. Gigerenzer and Murray proclaim that Kahneman and Tversky (1987, p. 155) have advocated this claim in the 1972 paper 'Subjective Probability: A Judgment of Representativeness', where it written that (Kahneman and Tversky 1972B, p. 451):

'Although our experimental examples were confined to well-defined sampling processes (where objective probability is readily computable), we conjecture that the same heuristic plays an important role in the evaluation of uncertainty in essentially unique situations where no "correct" answer is available.'

One might very well take this passage as a defense of claim b'2, such that it opens up the role of the representativeness heuristic in problems for which no unique solution exists, such as in the Tom W. study. Gigerenzer and Murray (1987, p. 155) however take it as evidence for the claim that judgments by representativeness are independent of content. It is not so clear what Gigerenzer and Murray mean with independence of content. It might entail at least one of the two following claims:

c3: People use a representativeness heuristic for all contents in the evaluation of uncertainty
c'3: The role of the representativeness heuristic in judgments of uncertainty is content-independent

It would have been easier to defend claim c3 or c'3 in case claim 2 'giving judgments by representativeness equals committing base-rate neglect', would hold. In that case, base-rate neglect would always follow from representativeness, independent of content. In section 3.2.3, claim 2 has been refuted though and this has repercussions for claims c3 and c'3 as I will come to explain.

Claim c'3 seems to be implicit in the use of 'however' in the following sentence in Gigerenzer and Murray (1987): 'However, in the recent past it has become more and more evident that if the structure of the problem posed is held constant, but the content is changed, responses differ' (pp. 154-155). Obviously, this phenomenon has been observed by Tversky and Kahneman themselves in the different versions of the Suicide Problem and the Taxicab Problem, where a change in content within an untouched Bayesian structure led to different results. Thus, if Gigerenzer and Murray were to defend claim c'3, this would be entirely misplaced.

Like claim c'3, claim c3 also lacks support. Recall that in the first version of the Suicide Problem, the representativeness heuristic is only applied to the conditional likelihoods and not to the base rates. In the second version however, a judgment by representativeness is also applied to the base rates. People do not use a representativeness heuristic for all contents. For some they do, for others not, dependent on how they evaluate the relation *'X is (very, …, not at all) representative of M'* as explained in section 3.4.1.

Gigerenzer and Murray (1987) admit that their claim concerning content cannot be upheld:

Kahneman and Tversky (1982[B]) also made a retreat, realizing that content is crucial and that, consequently, human reasoning cannot be described by content-independent formal rules. However, what has not been clarified is that judgments by representativeness—because they are in essence judgments by likelihood—are as content independent as formal Bayesian reasoning. (p. 156)

The content independency of judgments by representativeness has been refuted by the Suicide Problem. More generally, in section 3.4.1 it was explained that representativeness is taken by Tversky and Kahneman to be defined empirically by the subjects who, as repeatedly been shown, do take differences in content into account. In conclusion, Gigerenzer and Murray lack support for attaching the Content Insensitivity Attribute to the representativeness heuristic.

3.5.5. The Context Insensitivity Attribute

The Context Insensitivity Attribute describes that the representativeness heuristic, like Bayes' Theorem, operates independent of context. Like Bayes' Theorem, representativeness is a 'general-purpose' heuristic (Gilovic & Griffin 2002, p. 3). Embracing the representativeness heuristic Gigerenzer and Murray (1987) argue, therefore implies that 'the context of the problem is not very important for a theory of thinking' (p. 156). Furthermore, they argue that the representativeness heuristic 'seduces one into believing that thinking is directed by a few universal laws of thought' instead of the specific context (Gigerenzer and Murray 1987, pp. 156-157).

Gigerenzer and Murray (1987, p. 156) argue that the impact of contextual variables such as the wording of the problem and the particular example given are likely to be overlooked by proponents of the representativeness heuristic. For example, in the Engineer-Lawyer Study subjects are promised a bonus when their estimates come close to the highly accurate estimates of an expert panel. Given the accuracy of the experts, the subjects may now come to think that the personality descriptions are highly informative for those who are able to read them. Consequently, the subjects might come to understand that the best strategy to secure the bonus is to neglect the base rates and focus on the descriptions solely. Highly speculative as this possibility is, it can easily be checked for when rerunning the experiment without referring to the panel of experts as Gigerenzer and Murray add. Unfortunately Gigerenzer, Hell and Blank (1988) do not present different versions on the bonus in the discussion of the Engineer-Lawyer experiments they performed.

Other contextual variables that Gigerenzer has identified include visual versus verbal random sampling and the artificiality of personality descriptions as studied by Gigerenzer, Hell and Blank (1988) and discussed in section 2.3.2. Moreover, Gigerenzer, Hell and Blank (1988) present evidence that subjects have more ease in applying Bayesian statistical thinking when they are more familiar with the subject-matter. Specifically, subjects showed better results in predicting the outcomes of soccer matches than in the Engineer-Lawyer experiment. Finally, as discussed in section 2.4.4, Gigerenzer and Murray (1987) criticized the Taxicab Problem for ignoring the context of the example and underspecification of the base rates as a result.

The contextual variable pressed most on by Gigerenzer concerns information representation and frequency formats specifically. Reflecting on the debate on the representativeness heuristic, Gigerenzer (1993) advises us to look beyond mental algorithms. Like Arabic numbers are more suited than Roman numbers to the mathematical algorithm of multiplication, the specific representation of information is also key to the algorithm for intuitive statistics that people endorse. Indeed, Bayesian reasoning improves with frequency formats, though exceptions apply (Kahneman and Tversky 1996, p. 584-585).

Gigerenzer and Hoffrage (1995) add to this that it is not only the information format, but also the specific information menu that matters for people's performance on Bayesian problems. Take for example the Hammerton problem on medical screening. We are told that $P(H) = 0.01$, $P(E|H) = 0.9$, $P(\neg H) = 0.99$ and $P(E|\neg H) = 0.01$.[28] Gigerenzer and Hoffrage call this the standard information menu. In the short information menu, the following information is given: $P(H) * P(E|H) = 0.009$ and $P(\neg H) * P(E|\neg H) = .0099$. Especially when combined with frequency formats, the short information menu improves Bayesian reasoning.

As discussed in section 1.4.3, Kahneman and Tversky (1982A) have admitted that contextual variables such as the wording of a problem did not receive sufficient attention in their early studies. In that sense, they might agree with Gigerenzer and Murray (1987) that the impact of context has been underrated in their early experiments. At the same time, it should be said that exactly because Kahneman and Tversky (1982A) have identified this shortcoming, they have been able to take account of contextual variables.

Similar to the content insensitivity attribute, Kahneman and Tversky (1996, p. 583) have responded to the criticism of Gigerenzer and Murray (1987) regarding context as a misrepresentation of their position. In their defense, they indeed refer to Kahneman and Tversky (1982A). In this paper, Tversky and Kahneman discuss methodological considerations concerning the study of statistical intuitions. Among other issues, the role of context and the question-answering paradigm are discussed. Kahneman and Tversky (1982A, p. 129) argue that the research on judgment under uncertainty has shown that many adults do not have generally valid intuitions concerning the role of base rates in Bayesian inference. They stress however that it does not follow that any problem to which Bayesian inference can be applied, is answered incorrectly. In particular contexts, the rules of inference do seem to be compelling.

---

[28] With *H* for hypothesis and *E* for evidence. Recall that $P(H|E) = \frac{P(H)*P(E|H)}{P(H)*P(E|H)+P(\neg H)*P(E|\neg H)}$ [Eq. 7]

Examples of such contexts are given by Tversky and Kahneman (1982B) in their article 'Evidential Impact of Base Rates' and include the Taxicab Problem with causal base rates and the Intercom Problem. Although Gigerenzer and Tversky and Kahneman might argue about the validity of the normative Bayesian estimate for the problems, they seem to agree on the fact that the subjects' estimates are influenced by the particular context of the problem. Other examples provided by Tversky and Kahneman (1982B) are taken from Social Psychology. In a study by Ajzen (1977) subjects are told that 75% of class failed an exam. From this information, the subjects concluded that the exam was difficult. In a study on helping behavior with the same formal structure by Nisbett and Borgida (1975), subjects are told that most people do not give aid in some particular set-up. When confronted with examples, subjects however inferred that the participants in the study were mostly unfeeling brutes. Thus, they failed to recognize that the particular set-up makes people unhelpful. Tversky and Kahneman (1982B, p. 160) conclude that 'the major conclusion of this research is that the use or neglect of consensus information in individual prediction depends critically on the interpretation of that information.'

In defense of Tversky and Kahneman it should once more be emphasized that representativeness is defined empirically. That entails 'the relation "X is (very, …, not at all) representative of M" can be meaningfully assessed by judges' (Kahneman and Tversky 1982B, p. 88). There is no reason why judges would not take context into account in their assessment. It appears that some contexts, such as presented the Tom W. and Engineer-Lawyer studies, are very conducive to judgments of and by representativeness. Other contexts, such as Gigerenzer, Hell and Blank's soccer problem, are less likely to be associated with the representativeness heuristic. That Tversky and Kahneman have been more interested in focusing on those contexts in which judgments of representativeness are likely to occur, does not mean that no contexts exist in which these will not occur. Put differently, the representativeness heuristic does not operate independently of context. As a result, it is not implied by the heuristic that the context of a problem is not important for a theory of thinking.

To further prove the point made in last paragraph further, I would once more like to go back to 'On the Study of Statistical Intuitions'. Kahneman and Tversky (1982A) here argue that the 'properties that make formally equivalent problems easy or hard to solve appear to be related to the mental models, or schemas, that the problems evoke' (p. 130). With a reference to Hayes and Simon (1977), Tversky and Kahneman claim the reasoning process is schema- or content-bound. As a result, the specific context does make a difference for the operations or inference rules that are available. At this point, Tversky and Kahneman come to the conclusion that human reasoning cannot be sufficiently explained by use of content-independent rules.[29]

Immediately after their discussion of context- and content-sensitivity, Kahneman and Tversky (1982A) introduce the problems concerning Socratic instructions, as elaborated on in section 1.4.3. Although they do not explicitly mention this, one is led to think that Kahneman and Tversky (1982A) argue that the issue of Socratic hinting is also dependent on context. As a piece of information is interpreted differently depending on context, it might act as a Socratic hint one situation but not in another. Not without reason, Kahneman and Tversky (1982A) started the paragraph on Socratic hints with saying that 'the problem of mapping statistical or logical intuitions is *further* complicated' by it (p. 130).[30]

[29] As acknowledged by Gigerenzer and Murray (1987, p.156)
[30] Italics are not in the original text.

I conclude that the context insensitivity attribute cannot be defended. Although Gigerenzer has pointed out many relevant contextual variables that were initially overlooked by Tversky and Kahneman, it does not follow that Tversky and Kahneman in embracing the representativeness heuristic endorse a theory of thinking that is independent of context. On the contrary, Tversky and Kahneman have been very much concerned about the effect of contextual variables on people's judgments, were it only for the fact that they have studied the use of base rates in a wide variety of contexts. As an example, one might consult section 2.4, where I have extensively described the attention that researchers in the heuristics and biases program have for contextual details in the various versions of the Taxicab, Intercom, Suicide and Medical Screening problems.

3.6 Conclusion

In this chapter I have been concerned with the representativeness heuristic. I have tackled Gigerenzer's claim that this heuristic is only vaguely defined. More or less in line with Gigerenzer and Murray (1987), I have argued that judgments of representativeness can be described in terms of conditional likelihood. Moreover, I have criticized Gigerenzer and Murray (1987) for not carefully distinguishing between judgments of and by representativeness. I have shown that when this distinction is properly applied, Gigerenzer and Murray's claim that representativeness is a restatement of base-neglect cannot be upheld. What I have no objection against is the weaker claim that all cases of base-rate neglect can be described in terms of representativeness.

The failure to distinguish between judgments of and by representativeness has also made Gigerenzer overlook the explanatory role the representativeness heuristic has towards the apparent contrast in results to the experimental studies on statistical intuitions of the 1960's and those of the 1970's. Thus, contrary to what Gigerenzer (1996) has argued, the representativeness heuristic does have an explanatory role.

I have also addressed the concern that the representativeness heuristic is insufficiently limited in scope and does not give falsifiable predictions. I have shown that it does give falsifiable predictions. Moreover, I have argued that it is the subject who decides in the end whether a relation shows representativeness. In this decision, the subject also takes the context and content of a problem into account. Moreover I have shown that the contrast that Gigerenzer (1991A) sketches between representativeness and mental models is unfounded. I have furthermore claimed that the criticism of Gigerenzer (1996) that the representativeness heuristic lacks references to mental processes does not hurt Tversky and Kahneman's position, as the latter do not think that process models were fruitful at that time. I have argued that Tversky and Kahneman and Gigerenzer uphold different desiderata for psychological theory and that it is mostly an empirical matter which desiderata are most successful in attaining the shared goal of understanding the cognitive processes that produce both valid and invalid judgments.

Finally, I have studied whether the alleged attributes of Bayesian thinking are inherited by the representativeness heuristic. It appears that only for the inheritance of the Rationality Attribute a strong case can be made. I have argued that the Search Neglect Attribute and the Content and Context Insensitivity Attributes do not seem to apply to the representativeness heuristic as envisaged by Tversky and Kahneman.

All in all, most of Gigerenzer's criticism on the representativeness heuristic seems to be either misplaced or exaggerated. Because a lot of Gigerenzer's criticism on the representativeness heuristic cannot be substantiated, it does not hit Tversky and Kahneman as much. It is not surprising in this respect that Tversky and Kahneman (1996, p. 584) have commented that 'the refutation of a caricature can be no more than a caricature of refutation'. As a result, there is not much reason to take such a caricature seriously.

The potential disregard of Tversky and Kahneman towards him is a pity for Gigerenzer, as some of his criticism does make sense. For example, as shown in the Taxicab Problem, context and content are sometimes given insufficient attention by researchers in the heuristics and biases program. Moreover, neglect of search for information and a subsequent lack of external validity is an issue for some experiments where most information is already given. Furthermore, Gigerenzer does seem to have been right in pointing out that the heuristics and biases program has overemphasized rationality and has ignored the passions. On top of that, Gigerenzer's formulation of process models has led him to conduct new experiments which have enriched the study of statistical intuitions. Certainly in that sense, reliance on heuristics or process theories does not have to incompatible.

In his postscript, Gigerenzer (1996) has argued that the important issue dividing him from Tversky and Kahneman concerns research strategy. It is not so clear-cut however, that the difference in research strategy creates a gulf between the two camps. Tversky and Kahneman (1996) still maintain the goal to understand the cognitive processes underlying judgments. They are certainly not opposing the description of mental models. Tversky and Kahneman (1996) simply think it is not a fruitful approach. This does not mean at all that Tversky and Kahneman are in disagreement with Gigerenzer. One may wonder as a consequence, whether the divide in research strategy is that substantive.

I will now discuss how my thesis relates to the literature reviewed in section 1.2. Firstly, Samuels, Stich and Bishop (2002) have argued that the vagueness of the representativeness heuristic forms a real dispute between Gigerenzer and Tversky and Kahneman. It must be clear by now that Gigerenzer and Tversky and Kahneman are not in agreement when it comes the representativeness heuristic. At the same time though, the differences seem to come down to a supposed difference in research strategy. As a result, one can doubt whether one should label that as 'a very real dispute' as Samuels, Stich and Bishop (2002) do.

Secondly, my contribution is not incompatible with Jullien and Vallois' (2014) account that the differences between the two camps are driven by the objective-frequentist view endorsed by Gigerenzer and the subjective view of probability endorsed by Tversky and Kahneman. Their analysis could in principle also be extended to the representativeness heuristic, which according to Gigerenzer is closely related to Tversky and Kahneman's probabilistic views. Tversky and Kahneman have been fine with subjective models, definitions and heuristics, whereas Gigerenzer favors more objective process models. Similar to the point about norms and context discussed in section 2.5, this issue may make sense from a theoretical point of view. From a practical point of view though it might be less interesting, as Tversky and Kahneman themselves have purported that people think along the lines of mental models. In that sense, their position might be closer to Gigerenzer's than Jullien and Vallois picture it.

4. Final Remarks

In this thesis I have evaluated two points criticism cast by Gigerenzer towards the base-rate neglect studies conducted by the heuristics and biases program. The first point of criticism concerned the applicability of Bayes' theorem to the judgmental problems designed for these studies. The second point of criticism concerned the vagueness of the representativeness heuristic, which is supposed to explain the phenomenon of base-rate neglect. These two points of criticism are taken by Gigerenzer (1996) to be the 'main obstacles' between him and Tversky and Kahneman towards the shared goal of understanding the cognitive process underlying valid and invalid judgments.

Concerning the first claim I have firstly shown that deciding whether a judgment is valid or invalid is not a clear-cut issue. It is therefore difficult to say whether Gigerenzer’s criticism is mere pedantry or really substantial. It should be clear though, that both he and Tversky and Kahneman have not been sufficiently careful at times in checking the conditions for the proper application of Bayes’ theorem. I have argued that both parties have been eager to improve the design of the judgmental problems used in experiments. In that light, Gigerenzer’s criticism towards the applicability of Bayes’ theorem is mostly constructive. An undoubtedly constructive outcome of Gigerenzer’s criticism are the enriching empirical studies that he has conducted as a result of his criticism.

I have shown that Gigerenzer has been very critical of the representativeness heuristic. Most of his bold criticism towards the representativeness heuristic cannot be substantiated though, mostly for the reason that he has misrepresented the representativeness heuristic. For this reason, most of Gigerenzer’s criticism on the representativeness heuristic does not really hurt Tversky and Kahneman. In a sense, this should be pitied by Gigerenzer, as some of his criticism towards the representativeness heuristic did make sense, as Kahneman and Frederick (2002) also implicitly seem to argue in their revisionary article of the representativeness heuristic. It seems that Gigerenzer (1996) understood that most of his criticism failed to hit Tversky and Kahneman, as he retreated to a less aggressive but at the same time more dissociative position by claiming that he and Tversky and Kahneman are pursuing different research strategies. I have argued that it there is no fundamental reason though that would hinder a possible convergence of these strategies in the future. For a more comprehensive conclusion to Gigerenzer’s claims, I wish to refer the reader to sections 2.5 and 3.6.

Because the scope of this thesis is limited, it follows that limitations apply to the conclusions presented here. It cannot be assumed that the discussion between Gigerenzer and the heuristics and biases program on other issues than base-rate neglect can be evaluated in the same ways as I have done here. Similarly, as my analysis has been mostly limited to publications before 1996, it cannot do justice to the development of the debate on base-rate neglect afterwards. Analysis of publications after 1996 and of other biases may fill these gaps in my analysis. Moreover, I have only scarcely focused on the impact of the debate discussed here on later works by Gigerenzer and researchers in the heuristics and biases program. An analysis of that kind would more clearly bring out the relevance of the findings of my thesis for future scientific inquiry. An example of this might be the impact the base-rate neglect studies have had on the formulation of dual process theories by Kahneman (Kahneman & Frederick 2002) and the subsequent criticism that Gigerenzer as cast on these (Kruglanski & Gigerenzer 2011). A second example concerns the debate on libertarian paternalism as already touched upon in the introduction. From a more empirical point of view, it would be furthermore worthwhile to study whether a possible unification is possible between Gigerenzer’s process models and the representativeness heuristic, or its successor, attribute substitution.

5. Bibliography

Ajzen, I. (1977). *Intuitive theories of events and the effects of base-rate information on prediction*. Journal of Personality & Social Psychology, 35(5), 303.

Bar-Hillel, M. (1980). *The base-rate fallacy in probability judgments*. Acta Psychologica, 44(3), 211-233.

Buturovic, Z., & Tasic, S. (2015). *Kahneman's Failed Revolution Against Economic Orthodoxy.* Critical Review, 27(2), 127-145.

Edwards, W. (1968) Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: Wiley

Fiedler, K. & von Sydow, M. (2015). Heuristics & Biases: Beyond Tversky & Kahneman's (1974) Judgment under Uncertainty. In M. Eysenk & D. Groome (Eds.), *Revisiting the Classical Studies* (pp. 146-161). London: Sage Publications

Gigerenzer, G. (1991A). *How to make cognitive illusions disappear: Beyond "heuristics and biases".* European review of social psychology, 2(1), 83-115.

Gigerenzer, G. (1991B). *From tools to theories: A heuristic of discovery in cognitive psychology*. Psychological review, 98(2), 254.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. Maktelow & D. Over (Eds.), *Rationality Psychological & Philosophical Perspectives.* London: Routledge.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (& vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129-161). Chichester: Wiley

Gigerenzer, G. (1996). *On narrow norms and vague heuristics: A reply to Kahneman & Tversky*. Psychological review, 103(3), 592-96.

Gigerenzer, G. (2008). *Rationality for mortals: Risk and rules of thumb*. Oxford: Oxford University Press.

Gigerenzer, G. (2010). *Personal reflections on theory and psychology*. Theory & Psychology, 20(6), 733-743.

Gigerenzer, G. (2015). *On the supposed evidence for libertarian paternalism*. Review of philosophy and psychology, 6(3), 361-383.

Gigerenzer, G., Hell, W., & Blank, H. (1988*). Presentation and content: The use of base rates as a continuous variable*. Journal of Experimental Psychology: Human Perception and Performance, 14(3), 513.

Gigerenzer, G., & Hoffrage, U. (1995). *How to improve Bayesian reasoning without instruction: frequency formats*. Psychological review, 102(4), 684.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gigerenzer, G., & Sturm, T. (2007). Tools= theories= data? On some circular dynamics in cognitive science. In M. Ash & T. Sturm (Eds.), *Psychology's Territories Historical & Contemporary Perspectives from Different Disciplines* (pp. 305-343). Mahwah, NJ: Erlbaum

Gigerenzer, G., Swijtink S., Porter, Z., Daston, T., Beatty, L., & Krüger, J. L.(1989). *The empire of chance: How probability changed science and everyday life*. New York: Cambridge University Press.

Gilovich, T., & Griffin, D. (2002). Introduction – Heuristics and Biases: Then and Now. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Hacking, I. (2001*). An introduction to probability and inductive logic*. New York: Cambridge University Press.

Hammerton, M. (1973). *A case of radical probability estimation*. Journal of Experimental Psychology, 101(2), 252.

Hayes, J. R. & Simon, H. A. (1977) *Psychological differences among problem isomorphs*. In N. Castelian, Jr., D. B. Pisoni, & G. R. Potts (*eds.) Cognitive theory*. Hillsdale, NJ: Erlbaum

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*(4), 275.

Heukelom, F. (2005). *Gigerenzer the Decided* (Vol. 2). Tinbergen Institute Discussion Paper, 113.

Heukelom, F. (2009). *Kahnemann and Tversky and the making of behavioral economics*. Tinbergen Institute Research Series (No. 455).

Kahneman, D. (2002). *Transcript from an interview with Professor Daniel Kahneman and Vernon Smith*. Nobelprize.org. Nobel Media AB 2014. Retrieved from: <http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2002/kahneman-interview-transcript.html>

Kahneman, D. (2003*). Maps of bounded rationality: Psychology for behavioral economics*. The American economic review, 93(5), 1449-1475.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 49-81). New York: Cambridge University Press.

Kahneman, D., Slovic, P., Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kahneman, D. & A. Tversky, (1972A). *On the Psychology of Prediction*. Oregon Research Institute Research Bulletin 12, 4.

Kahneman, D., & Tversky, A. (1972B). *Subjective probability: A judgment of representativeness*. Cognitive psychology, 3(3), 430-454.

Kahneman, D., & Tversky, A. (1973). *On the psychology of prediction*. Psychological review, 80(4), 237.

Kahneman, D., & Tversky, A. (1982A). *On the study of statistical intuitions*. Cognition, 11(2), 123-141.

Kahneman, D., & Tversky, A. (1982B). Judgments of and by representativeness. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics & biases* (pp. 84-98). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). *On the reality of cognitive illusions*. Psychological review, 103(3), 582-91.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. Psychological review, 118(1), 97.

Likelihood. (March 2016). In *OED Online*. Oxford University Press. Retrieved from: http://www.oed.com/view/Entry/108313?redirectedFrom=likelihood (accessed May 31, 2016).

Lyon, D., & Slovic, P. (1976). *Dominance of accuracy information and neglect of base rates in probability estimation*. Acta Psychologica, 40(4), 287-298.

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, *36*(9), 929.

Nisbett, R. E., & Borgida, E. (1975). *Attribution and the psychology of prediction*. Journal of Personality and Social Psychology, 32(5), 932.

Phillips, L. D., & Edwards, W. (1966). *Conservatism in a simple probability inference task*. Journal of experimental psychology, 72(3), 346

Rescorla, Michael (2015). The Computational Theory of Mind. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu/archives/win2015/entries/computational-mind/>.

Samuels, R., Stich, S. & Bishop M. (2002). Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear. In R. Elio (Ed.), *Common sense, reasoning, & rationality*. Oxford: Oxford University Press.

Shleifer, A. (2012). *Psychologists at the Gate: A Review of Daniel Kahneman's" Thinking, Fast and Slow"*. Journal of Economic Literature, 1080-1091.

Stein, E. (1996). *Without Good Reason*. Oxford: Clarendon Press

Thaler, R.H., & Sunstein, C.R. (2003). *Libertarian paternalism*. American Economic Review 93(2): 175–179.

Tversky, A., & Kahneman, D. (1971). *Belief in the law of small numbers*. Psychological bulletin, 76(2), 105-110.

Tversky, A., & Kahneman, D. (1974). *Judgment under uncertainty: Heuristics and biases.* Science, 185(4157), 1124-1131. T

Tversky, A., & Kahneman, D. (1977A). Causal thinking in judgment under uncertainty. In R. Butts & J. Hintekka (Eds.), *Basic problems in methodology and linguistics* (pp. 167-190). Dordrecht, Holl&: D. Reidel Publishing Company.

Tversky, A., & Kahneman, D. (1977B). *Causal Schemata in Judgments under Uncertainty*. Decisions & Designs, Incorporated.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 49-72). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1982A). Causal Schemata in judgments under uncertainty. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 117-128). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982B). Evidential impact of base rates. In D. Kahneman, P. Slovic & A. Tversky (Eds*.), Judgment under uncertainty: Heuristics and biases* (pp. 153-160). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). *Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment*. Psychological review, 90(4), 293.