Title

Re-Examining Base-Rate Neglect: The Effect Of Context

Permalink

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Adler, Nine

Dewitt, Stephen H

Publication Date

2024

Re-Examining Base-Rate Neglect: The Effect Of Context

Nine Adler[1], Stephen H. Dewitt[1]

[1]Department of Experimental Psychology, University College London, 26 Bedford Way, WC1H 0AP

Abstract

Classic base-rate neglect studies have been consistently criticised for lacking ecological validity. A study by Welsh & Navarro (2012) found this heuristic was significantly reduced when participants perceived the base rate as more relevant. The present study aims to study this phenomenon through a more realistic scenario while simultaneously capturing participants' written reasoning. Using mixed-methods, participants (N = 2,052) read an engaging scenario regarding a person who committed infidelity and containing a base-rate and specific information where the contextual information regarding the base-rate was manipulated. They were then asked to provide an estimate of the person's likelihood to cheat in the future. Results show that each of our three manipulations to the context of the base rate are significant in affecting participants' estimates, supporting Welsh and Navarro's findings. Analysis of participants' written reasoning demonstrates the sophistication and nuance of participants' engagement with the base-rate, challenging the original view of this supposed heuristic.

Keywords: causal reasoning; base-rate fallacy; ecological validity; heuristics; mixed-methods; intuitive theories

Introduction

Throughout the 70's and 80's Amos Tversky and Daniel Kahneman pioneered the heuristics and biases' field. They have coined and studied over a dozen biases, heuristics and reasoning fallacies, which seem to support the idea that the human mind is susceptible to deception and can be easily misled (Tversky & Kahneman, 1982).

Heuristics are mental shortcuts that individuals resort to in order to approximate an answer to a problem. Tversky & Kahneman (1982) describe heuristics as having "a logic of [their] own, which departs systematically from the logic of probability" (p.88). Typically, this has led to a vision of human reasoning as relatively 'crude', with Kahneman and Tversky (1972) describing their results as "discouraging", Nisbett et al. (1983) stating that the use of heuristics itself was "disturbing" and even that their findings had "bleak implications for human rationality" (Nisbett & Borgida, 1975, p. 935). Fiske and Taylor (1991) went as far as describing the human mind as a "cognitive miser".

Intuitive theories

However, a more recent body of work presents human reasoning as complex, sophisticated and subtle. Tenenbaum et al.'s (2007) 'intuitive theories' framework posits that humans draw from their extensive personal knowledge of how the world works to make sense of the limited and ambiguous data they have available. This has parallels with Klein et al.'s (2006) 'data/frame theory' where humans build frames – analogous to mental models – in which data is agglomerated. These frames are highly flexible and adaptable and go through endless iterations (where data is discovered, inferred, or even discarded), questioning (where data quality is evaluated) and reframing (where frames themselves are compared to one another). Figure 1 broadly represents the combined view of intuitive theories and data/frame theories.
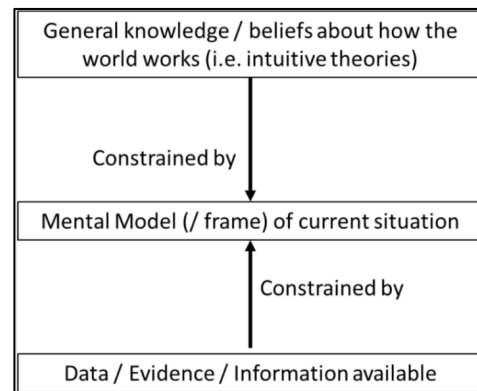

Figure 1. Combined representation of intuitive theories and data/frame theories

The present research aims to re-examine classic heuristics and biases from an intuitive theory / data-frame perspective. To this end, and in line with recent calls for more process-oriented studies (McNair, 2015; Tubau et al., 2015), we employ mixed methods, giving participants the opportunity to explain their reasoning which has often been lacking in this field.

Current study

This experiment focuses on a widely studied phenomenon called base-rate neglect (see Allen et al., 2006 for a meta-analysis). Base-rate neglect occurs when people tend to favour new data over established priors (Kahneman & Tversky, 1973). This goes against Bayesian inference, a fundamental principle of probability theory that requires the updating of prior probabilistic information with new probabilistic information, yielding posterior data (Laplace & Simon, 1951). One of the most famous experiments used to study base-rate neglect (BRN) is the taxi-cab problem by Bar-Hillel (1980). The problem is as follows:

Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent 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 later identified the cab as a Green cab.
The court tested the witness' ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found that the witness was able to identify each color correctly about 80% of the time, but confused it with the other color about 20% of the time. What do you think are the chances that the errant cab was indeed Green, as the witness claimed?

If one were to apply Bayes' theorem, the answer to this question would be 41%. Yet, on average, Bar-Hillel's respondents answered 80%, which is often taken to suggest they are focusing purely on the new specific piece of data (the witness's 80% accuracy), ignoring the base-rate about the overall "cab population".

However, Welsh and Navarro (2012) questioned the assumption that participants accept the base rate as the researchers intended and aimed to demonstrate people's ability to judge the trustworthiness of a base rate based on contextual factors. In their study, participants envision themselves as part of a research team identifying potential threats from predators in an unexplored location. Previous observations classified a certain percentage of predators as threats (the prior). Since then, new data with varying proportions has also been collected. Participants were tasked with estimating the proportion of threatening predators in the overall population, combining these two figures. To manipulate the trustworthiness of the base rate, the researchers altered key environmental factors across conditions: data collection location, data age, data source, and the number of observations. By manipulating these variables, they came up with 32 scenarios ranging on a spectrum of 'low trust' to 'high trust'. They found that the scenarios where the base rate data were more trustworthy yielded lower levels of base-rate neglect.

We also aimed to manipulate contextual factors to see how these would affect our participants' reasoning, however we wished to also use a scenario closer to a typical real-life situation. We presented participants with an engaging scenario, using an emotionally potent topic – infidelity – to see whether context affected participants' use of the base-rate. In our scenario, the base-rate was that a person had cheated on one of three known ex-partners. The factors we manipulated were the source of the cheating confession (from the person themselves or from their ex-partner), the type of cheating (cheated with one person vs multiple people) and the recency of the cheating (the cheating happened while with the latest ex-partner or the oldest ex-partner).

In line with Welsh and Navarro, we hypothesised that all three independent variables would have a significant effect on participants' confidence in the partner's ability to stay faithful in the future, with higher estimates of the partner's future faithfulness when they reported it themselves (source), when it was a partner from further in the past (recency) and when they only cheated on that partner with a single person (type).

Methods

Participants

A total of 2,052 participants completed the study and were recruited from Prolific Academic. Participants ranged from 18 to 82 (M=41.7, SD=13.7). We required an equal distribution of gender (Female = 49.1%; Men = 49.2%; Other = 1.2%; Prefer not to say = 0.6%). Our inclusion criteria were that participants must be over 18, reside in the United Kingdom and be first language English speakers. They were compensated £9/h and on average, participants took 4.3 minutes to complete the study (SD = 3.3).

Design

The design was a 2x2x2 between participants study with participants being randomly assigned to one of the eight conditions.

Our three independent variables were loosely based on Welsh and Navarro's design but adapted to our scenario: source of reveal, recency of cheating and type of cheating each had two associated levels. Source of reveal was split between internal (the date confesses to have cheated) and external (the ex-partner reveals they were cheated on). Recency of cheating was split between recent (they cheated on their most recent ex) and old (they cheated on their oldest ex). Lastly, the type of cheating was split between multiple (they cheated on their ex-partner with multiple people) and one (they cheated on their ex-partner with one person).
Our dependent variable was participants' numerical input between 0 and 100 about how likely the partner in the scenario would be faithful.

Methods & Procedure

Participants began by giving their informed consent. Our scenario was presented as a vignette text block with the introduction, specific information, numerical and open questions being common across conditions. We chose

gender-neutral names and non-gendered pronouns – which we disclosed in the experiment's brief – so that participants could best represent themselves in the scenario, independently of their gender and sexual orientation.

On a new screen page, participants read the following introduction before clicking to move to the next stage of the experiment: "*You have been casually dating Sam for a few months and have been getting on well. You knew them a little bit before dating them through your extended social groups and are already acquainted with 3 of their ex-partners, but you don't know them well. You are due to meet Sam for dinner tonight. However, you've noticed they were a bit quiet last time you met up and you haven't heard from them for a few days.*".

Participants were then randomly allocated to one of our eight conditions which revealed that their date had previously cheated in one of their relationships. While the base-rate (1 out of 3) remained consistent across conditions, we manipulated the source of the reveal (from the date or from the ex that was cheated on), the recency of the cheating (either the previous relationship or their first relationship) and the magnitude of the cheating (cheated with one other person or multiple people).

All participants then saw the following contextual information regarding the cheating reveal: "*Throughout dinner you and Sam have a discussion about the cheating incident. During the later part of the dinner Sam tells you that they have strong feelings for you and would like to make the relationship official between you both and says that they've been a bit nervous about asking you. They tell you not to answer now but take your time to think.*". They then saw the following specific information: "*You decide to get the opinion of three of your closest friends who also know Sam a bit and you tell them everything. Two of them feel like Sam has changed since then, but one isn't so sure.*".

The information they had read before was repeated in a text box with a smaller font, to serve as a reminder. Two questions followed, one quantitative and one qualitative. For the quantitative question, participants had to place a cursor on a slider ranging from 0 ("Completely unconfident") to 100 ("Completely confident"). The cursor was placed in the middle by default and participants could not move to the next part of the experiment before moving or clicking on the cursor (if they wanted to select 50, they would have to move the cursor elsewhere and then reposition it in the middle). We then presented the qualitative question with an open text box asking participants to explain their previous response in as much detail as possible.

Results

Descriptive statistics

In terms of participants' confidence in Sam's faithfulness, the mean confidence in the "Internal/Older/One" condition was the highest ($M = 55.11$; $SD = 24.39$), compared to the "Internal/Recent/Multiple" condition, which was the lowest

($M = 38.95$; $SD = 25.59$). Table 1 shows a summary of these results.

Table 1. Summary of quantitative results

	Condition	Mean rating (%)	SD (%)
Source	Internal	48.1	25.8
	External	50.6	25.8
Recency	Recent	47.1	25.8
	Older	51.5	27.5
Type	One	53.1	25.6
	Multiple	47.7	25.6

Inferential statistics

Firstly, we had concerns about observing effects of age and gender due to the nature of our scenario. We ran a linear regression between age of participants and their answer to the numerical question and found a non-significant relationship ($F(66, 1985) = 1.27$, $p = .074$). We then did the same with gender and once again found a non-significant relationship ($F(3, 2048) = 1.46$, $p = .225$).

We then ran an ANOVA with three independent variables – source, recency, and type of cheating – and one dependent variable, participants' belief of Sam's faithfulness. We observed a significant result for our Source variable ($F(1, 2045) = 5.77$, $p = .016$), where participants who found out about the cheating through the ex-partner thought Sam was more likely to be faithful ($M_{external} = 50.6$; $M_{internal} = 48.1$).

We also found a significant effect of our recency variable ($F(1, 2045) = 15.31$, $p < .001$) where participants learning that Sam had cheated on their oldest ex-partner saw them as more likely to be faithful than those finding out Sam cheated on their most recent ex-partner ($M_{older} = 51.5$; $M_{recent} = 47.1$).

Finally, the type of cheating was also a significant factor ($F(1, 2045) = 43.01$, $p < .001$), where Sam cheating on the ex-partner with one person led to higher ratings about Sam's ability to be faithful compared to when Sam cheated on their ex-partner with multiple people ($M_{one} = 53.1$ $M_{multiple} = 45.7$). All three results show that our manipulation to the contextual information regarding the base-rate led to significant differences in our participants' responses, which is line with our main hypothesis, except for our source variable which did not go in our predicted direction. Figure 2 depicts these results as a line graph.
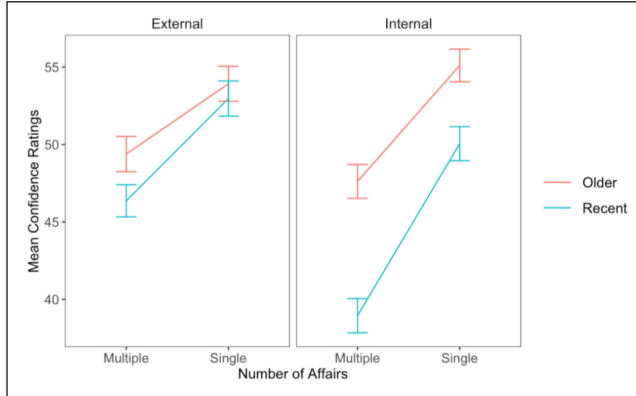
Figure 2. Line graph with mean confidence ratings (as %) per condition

We also found a significant two-way interaction between our source and recency variables ($F(1, 2045) = 4.73$, $p = .03$), which may suggest that when participants learned Sam cheated on their most recent ex-partner, finding out through Sam or the ex-partner led to the greatest difference in confidence ratings about Sam's likelihood to be faithful. We did not find any other significant interactions.

Qualitative data

Following familiarisation with the data, we developed a coding scheme driven by our research question of whether participants' judgements were influenced by the context which varied between conditions, so our primary codes aimed to document instances in which participants mentioned the key condition manipulations (e.g., source, recency, and type of cheating). For each of these we coded whether these factors were mentioned positively (in favour of Sam's trustworthiness), negatively, neutrally or were dismissed. We also coded whether participants mentioned the base-rate and specific information in the same way as above. Finally, we also noticed a particular heuristic of interest which we coded for and will be discussed further below.

Any data not fitting into the above codes, showing misunderstanding of the scenario, or was simply uninterpretable was coded as "Unclassified" as well. Overall, nearly 40% of responses were coded as such ($N = 815$). A summary of the data can be seen in Table 2, and we will discuss the results separately, code by code below.

Table 2. Coding results in % (Pos = positive, Neg = negative, Neu = neutral, Dis = dismissed, BR = base-rate, SI = specific information, H = heuristic, S = Source, R = recency, T = type)

		Pos	Neg	Neu	Dis	N	% total
BR		37.2	48.7	7.7	6.4	2052	3.8
SI		53.5	21.5	11.9	13.1	2052	15.2
H		100	-	-	-	2052	16.5
S	Int	85.7	4.8	8.6	0.9	1007	33.4
	Ext	43	39.7	23.3	19.8	1046	11.1
R	Rec	0	78	20.3	1.7	1001	5.9
	Old	84.1	2.9	13	0	1052	6.6
T	One	87.5	0	12.5	0	1011	0.8
	Mul	0	97.2	1.7	1.1	1042	17.3

Base rate. The base-rate was the least mentioned code, likely due to our conservative coding approach: participants had to explicitly mention that Sam had cheated on one out of three previous exes to be assigned this code. If we had coded anyone who mentioned the cheating incident almost all participants would have received this code. This becomes apparent when we see that it is mostly mentioned positively (46.2%), to emphasise that Sam only cheated on a third of their partners, where the base-rate therefore needs to be put in perspective: "she only cheated on 1 of her ex boyfriends not all of them" (P.1896). When coded negatively, we see the base-rate is approached with caution "Having already had at least three partners and cheated on a least one of them" (P.89), showing not only general doubt in the base-rate itself but the acknowledgment that the base-rate only applies to a sub-sample of Sam's partners.

Specific information. This code was assigned when participants mentioned that two out of the three friends thought Sam had changed. Unsurprisingly, when mentioned positively, participants focused on the two out of three friends who believed Sam had changed, "If 2 people that I am close with believe that Sam has changed then I would put my faith in them" (P.1565). On the other hand, when mentioned negatively, emphasis was put on the remaining unsure friend, "I also feel like if any one of my friends thinks that someone I'm dating is bad news then I should take this into consideration" (P.96). Some participants actively dismissed the specific information, either because the base-rate itself made it irrelevant "Despite the two people saying they think Sam has "changed", actions speak louder than words" (P.1754), or because it didn't seem strong enough "I didn't really take my friends' opinions into account as they only know him "a bit"" (P.83).

Source of reveal. 'Internal' was coded for if participants in that condition mentioned the fact that Sam told them about the cheating incident. Similarly for external, participants in that condition had to state the ex-partner told them about the cheating incident, or that Sam didn't tell them themself. Nearly half of our sample (44.5%) mentioned the source of the reveal in the answer, with people in the internal condition raising it nearly three times more than those in the external condition (33.4% and 11.1% respectively). Almost unanimously, Sam revealing the cheating was seen positively (85.7%), as a sign of being "upfront" (P.1641), "genuine" (P.1199), "honest without being prompted" (P.1257), a sign of "how much they have changed" (P.1378). Alternatively, a

minority of people (4.8%) were put off by it entirely "The fact that Sam felt the need to discuss former partners and confess infidelity to me suggests that s/he has not yet 'moved on' and makes me question just how prepared s/he is for another relationship." (P.1936).

When the source of reveal was external, participants were much more divided. The majority saw it positively (43%), "even Alex says she thinks he regrets it" (P.503), seeing Alex's words as solid evidence: "an ex is prepared to vouch for him" (P.1240). Those who did not see it in such good light had various reasons to do so. Some wish they had found out through Sam instead "I would also be slightly worried that Sam did not appear to tell me about this until her ex mentioned it" (P.86). Nonetheless, some participants presented an opposite reasoning by paying close attention to the wording "Alex only says they think Sam feels bad about this. Which reads like Alex is inferring it from Sam's behaviour rather than any concrete evidence" (P.172). Finally, there was speculation about Alex's motive in revealing they had been cheated on: "an ex who was cheated on may very well have their own agenda to stir things up" (P.1925).

Recency of cheating. Participants in the "Recent" condition had to mention the fact that Sam cheated on their most recent partner and contrastingly, participants in the "Older" condition had to mention the fact that Sam cheated on their oldest partner. The recency of the cheating was the least mentioned factor out of the three manipulations (12.5%). When it was brought up, there seemed to be a consensus, when the cheating happened while with the most recent partner, it was mainly negative (78%) with participants doubting Sam could have changed in such a short period of time "Well the girl he cheated on was his most recent relationship so he has done it not that long ago seemingly which would make me doubt whether he would be faithful to me" (P.31). Contrastingly, when the cheating happened with the oldest partner, it seemed more forgivable "They were likely a lot younger" (P.1378), "Perhaps now they may have matured" (P.387).

Type of cheating. Participants in the 'One' condition had to write that Sam cheated with a single person in order to get that code whereas participants in the 'Multiple' condition had to write that Sam cheated with multiple people. Although the type of cheating was only mentioned by less than a fifth of participants (18.1%), nearly all mentions come from the multiple condition (17.3% of total participants). These were also mainly negative (97.2%), with participants raising concerns about Sam's ability to be faithful in the future "I do believe people can make stupid mistakes but if you've done it more than once, it's a very telling sign" (P.1730).

Heuristic. We observed a type of reasoning that could be described as a heuristic, using very categorical thinking (e.g., 'Once a cheater always a cheater'), and marked it as absent or present. We only considered answers that encompassed an

'all or nothing' mindset and conveyed certainty that Sam could not be trusted ever again and would definitely cheat. For this code, we observed multiple variations of categorical thinking with popular sayings such as "Once a cheat, always a cheat!" (P.1238) or "A leopard doesn't change its spots" (P.192). Some went into more detail yet keeping a very strict mindset "I am of the opinion that there is never a good reason to cheat on a partner" (P.1773).

Participants who were coded for this reasoning were starkly pessimistic about Sam's likelihood to remain faithful ($M = 25.3$, $SD = 25.8$) compared to all other participants ($M = 56$, $SD = 25.8$). A t-test revealed this difference was significant ($t(2050) = 22.2$, $p < .0001$) which suggests consistency between participants qualitative and quantitative answers.

Additionally, participants had significantly shorter responses when using this type of thinking ($t(2050) = 8.05$, $p < .0001$) using on average 13 less words than participants who did not display similar heuristic-like thinking ($M_{heuristics} = 33.6$, $SD_{heuristics} = 30.05$; $M_{non-heuristics} = 46.6$, $SD_{non-heuristic} = 30.2$). This could be interpreted as evidence that participants use this type of thinking as a literal short-cut, to shorten their thinking and subsequently, its explanation. However, we might suppose that because this reasoning lacks nuance it is naturally more concise.

Discussion

This study aimed to look at whether, in a base rate scenario, people's use of that statistic is heuristic-like (e.g. showing simple BRN) or affected by context. More generally we aimed to see if participants' reasoning, particularly in the qualitative data bears a resemblance to the intuitive theories / data-frame models. We hypothesised that our three variables – source of reveal of cheating, recency of cheating and type of cheating – would affect participants' belief that our character would be faithful in their next relationship, and by extension, their 'use' of the base-rate. We found all three parameters to be significant predictors, in line with Welsh and Navarro's study (2012) about how context affects the interpretation of base-rates. We also collected qualitative data to have a better understanding of participants' reasoning which allows us to have a more nuanced and complete picture of the underlying processes at play.

Only a minority of participants were coded as mentioning the base-rate in their written responses. This is because we chose a highly conservative approach to coding for the base-rate, requiring participants to explicitly mention it in a numerical format. It is therefore not a good indicator of base rate neglect. Since the cheating incident was so central the vast majority of participants that did not fit this code still mentioned the cheating incident but simply did not express this numerically.

In their qualitative data, a large portion of our sample mentioned the contextual factors while also drawing on their personal knowledge of how people tend to behave to build an impression of Sam's character and infer their trustworthiness

from that impression. We saw many displays of nuanced thinking when considering what kinds of cheating behaviour tend to be more or less indicative of trustworthiness, with participants typically seeing more recent cheating, cheating with multiple people, and Sam not being the one to tell us about the incident as more indicative of current bad character and low trustworthiness, in line with the quantitative findings. Overall, our qualitative data bears a strong resemblance to the intuitive theory / data-frame model of human reasoning and does not on the whole resemble simple heuristic-like thinking. Interestingly however, a substantial portion of participants were coded as presenting a particular 'heuristic'-like form of thinking. This was not base rate neglect per se, but neglect of all information other than that Sam had cheated: a common phrase was 'once a cheater, always a cheater' suggesting that regardless of number of time or any other factors, if someone cheats, they are completely untrustworthy. It is worth noting that this may well still be in line with intuitive theories as this reasoning may very well stem from personal history or experiences that have led to the adoption of this belief which extend beyond cheating entirely. Some participants mentioned personality traits as the main reason for their answer "I am naturally a person who finds it difficult to trust others even if I have no reason to doubt them" (P.664). Indeed, frequent observations of persistent negative behaviour may tint an individual's global outlook and therefore the simple idea that people will continue to behave like they have in the past can then be applied to scenarios the person has not (yet) experienced themselves.

It is questionable whether our base-rate is comparable to those used in previous studies (Bar-Hillel, 1980; Cosmides & Tooby, 1996; Gigerenzer, 1996; Kahneman & Tversky, 1996) as it did not rely on a population statistic. Instead, our base-rate concerned the history of the individual whom participants were asked to make a judgement about, which may rightly be considered more 'relevant' than a population base-rate (e.g., the proportion of people who cheat). However, previous studies have in fact been inconsistent in their definition of base-rate, with some simply equating it with a prior (Ajzen,1977), which would encompass the base-rate in our study, while others have linked it more tightly to a statistic regarding a population (Klayman & Ha, 1987).Yet it is important to note that there is no general consensus on what exactly qualifies as a base-rate, with Welsh & Navarro (2012) stating that "every real world base-rate is potentially misleading" (p.12). This perhaps reinforces the idea that the original studies about base-rate neglect might not necessarily be studying a 'real world' phenomenon (McKenzie, 2003) and that we need to study more naturalistic situations with a greater variety of forms of prior. While we did aim for our study to be comparable to those such as the taxi-cab problem and its numerous variants, our problem was not a Bayesian updating task, there was no computation to perform, no normative answer to reach and the likelihood participants had to come up with was about a future event. We argue this also proposes a novel way of studying base-rate neglect without

solution relying on mathematical numeracy and therefore potentially discriminating between participants' personal abilities in this regard (Bruckmaier et al., 2021). Few real-life decisions are mathematically solvable and so there is an inevitable trade-off between ecological validity and adherence to previous paradigms. While we wished in this work to be able to compare our findings to previous work, we simultaneously wished to develop a scenario which felt more like a real-world situation than classic studies.

When opting for a topic as emotional as infidelity (Sharpe et al. 2013), we must consider that participants will bring their own data to the table, which is precisely the type of response we wished to elicit. Similarly, participants who would describe themselves as "being a forgiving person" (P.889) might attribute very little weight to data relating to past events. This concords with McKenzie's (2003) conclusion about how researchers should not view strategies devised for laboratory settings and everyday life as similar or even comparable. In problems such as the taxicab problem, it is difficult to imagine participants feeling emotionally and personally invested in the story and using past experiences to reach their decision.

Fundamentally, this experiment lays the groundwork for future studies wishing to investigate causal reasoning and heuristic thinking while maintaining high ecological validity. It offers a rich and highly customisable methodology and a flexible qualitative analysis to better accumulate evidence of sophisticated and nuanced thinking, as well as showing how people integrate information regardless of its diagnosticity.

Acknowledgments

References

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

Allen, M., Preiss, R. W., & Gayle, B. M. (2006). Meta-analytic examination of the base-rate fallacy. *Communication Research Reports*, 23(1), 45-51.

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

Bruckmaier, G., Krauss, S., Binder, K., Hilbert, S., & Brunner, M. (2021). Tversky and Kahneman's cognitive illusions: who can solve them, and why?. *Frontiers in Psychology*, 12, 584689.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from

the literature on judgment under uncertainty. *cognition*, 58(1), 1-73.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. Mcgraw-Hill Book Company.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky.

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

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, 246(1), 160-173.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, 94(2), 211.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent systems*, 21(5), 88-92.

Laplace, P. S., & Simon, P. (1951). *A philosophical essay on probabilities*, translated from the 6th French edition by Frederick Wilson Truscott and Frederick Lincoln Emory.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of development Economics*, 99(2), 210-221.

McNair, S. J. (2015). Beyond the status-quo: research on Bayesian reasoning must develop in both theory and method. *Frontiers in Psychology*, 6, 97.

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

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological review*, 90(4), 339.

Sharpe, D. I., Walters, A. S., & Goren, M. J. (2013). Effect of cheating experience on attitudes toward infidelity. *Sexuality & Culture*, 17, 643-658.

Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. *Causal learning: Psychology, philosophy, and computation*, 301-322.

Tubau, E., Aguilar-Lleyda, D., & Johnson, E. D. (2015). Reasoning and choice in the Monty Hall Dilemma (MHD): implications for improving Bayesian reasoning. *Frontiers in psychology*, 6, 353.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

Welsh, M. B., & Navarro, D. J. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes*, 119(1), 1-14.