

Predicting Framed Decisions: Simulation or Theory?

Anton Kühberger^{1*}, Claudia Luger-Bazinger²

¹Department of Psychology & Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria

²Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens, Salzburg, Austria

Email: *anton.kuehberger@sbg.ac.at

Received 13 April 2016; accepted 24 June 2016; published 27 June 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

When predicting thoughts and behavior of other people, we use either the self as the basis for predictions (i.e., we simulate others), or theoretical knowledge (i.e., we use knowledge about others). To find out whether the prediction of complex choices is possible we asked participants to predict the choice of a well-known or unknown target person in the classic Asian disease framing task, a paradigmatic example of a paradoxical decision. In addition, we collected participants' self-reports on their prediction strategies (theory or simulation). People's choice in a framing task was correctly predicted regardless of familiarity. Although familiarity influenced reported strategy, correct predictions were presumably based on simulation, since explicit theoretical knowledge either was irrelevant or non-existent. These findings show that the correct prediction of paradoxical decisions by high-level simulation cannot be ruled as an alternative to prediction by theory.

Keywords

Framing, Rationality, Simulation-Theory, Theory-Theory

1. Introduction

Reading other people's minds is essential for our social life. Therefore, the question never was whether humans are able to read others' minds, but rather how they accomplish it. Two basic positions have been proposed: First, according to theory-theory (TT) we predict other people's thoughts and actions by applying a set of rules, which consists of naïve knowledge about the mental domain (i.e., a folk psychology). Thus, people possess a theory of mind encompassing a set of laws on how external stimuli become inner states (e.g., perceptions; world-to-mind);

*Corresponding author.

how some inner states (e.g., beliefs and desires) lead to other inner states (e.g., preferences; mind-to-mind); and how certain inner states (e.g., preferences) lead to behavior (e.g., choices; mind-to-world). Knowledge about prototypical cases, empirical generalizations, schemas, scripts, and narratives do all qualify as a “theory”. When applying a theory of mind other people’s behavior is explained and predicted by ascribing beliefs, desires, and other mental states to them. For instance, for predicting why Tom has opened the fridge, I could conclude by theory that Tom has opened the fridge because he was thirsty, he knew that juice was in the fridge, and had the goal to drink. This would be an example of applying the “belief-desire-law” which holds that people will do what satisfies their desires (Gauker, 2005). The theory theory’s view strongest advocates are some philosophers (Churchland, 1981; Fodor, 1987), and some developmental psychologists (e.g., Gopnik, 1993).

Alternatively, simulation theory (ST; Goldman, 2006; Gordon, 1986) states that predictions are not based on theories about the mental domain. Rather, we use our own mental mechanisms for predicting the mental processes of other people by having our own mental system running offline and producing the same sequence of mental processes that can be found in another person. Loosely speaking, a mental state, or process, is a simulation of another mental state, or process, if it imitates, copies, or reproduces the second event.

Mental simulation can be either intra-personal (e.g., when predicting own behavior in hypothetical situations), or inter-personal (e.g., when predicting the behavior of another person). Goldman (2006, 2009) distinguishes between two forms of mental simulation: ‘low-level simulation’ is mirroring (or resonance; Saxe, 2005); it is automatic, widely unconscious, and does not rely on task-specific knowledge or information. There is strong evidence that at least some basic emotions result in experiencing these emotions on observation of their facial expression (e.g., Goldman & Sripada, 2005). This is a simulation process, as the attribution of, say fear, of the observing person Y at person X is based on mirroring, and thus re-experiencing, X’s fear. In the neuroscience literature this account of simulation is also termed the “direct recruitment hypothesis”, as one simulates another person’s acts, including his or her internal states, as if it were one’s own internal states by employing the same neural circuitry that one uses for one’s own processes (Amodio & Frith, 2006).

“High-level simulation” is pretence-driven, or imagination-based, simulation. This form of simulation is effortful, and often conscious. It frequently involves complex mental states, is based on task-specific knowledge and information, and tends to involve propositional attitudes. The central process here is *imagination*: attempting to produce some mental state in oneself by endogenous means (Shanton & Goldman, 2010).

Unlike in the early days of the TT-ST debate, the dominant understanding now is of a “simulation-theory-hybrid” in which TT and ST do co-exist (Bach, 2011; Goldman, 2006; Heal, 1998; Malle & Hodges, 2005). Most notably, high-level simulation needs to be informed by some sort of theory about what the relevant features of the person, or situation, are that have to be imagined. This poses a threat of collapse (Perner & Brandl, 2009). Indeed, attempts to draw a clear distinction confront critique at the theoretical, empirical, and neuronal level (e.g., Bazinger & Kühberger, 2012a, 2012b; Ravenscroft, 2003; Saxe, 2005; Schurz et al., 2015).

TT is very flexible since it is not committed to any specific form of theory: Theories can be vague and inconsistent, and at the same time implicit and ineffable. Therefore, a mindreading task where no form of theory could be brought to bear is difficult to imagine (Perner & Kühberger, 2005). ST, on the other hand, is less flexible. The core features of any form of simulation are *resemblance* and *reuse* (Gallese & Sinigaglia, 2011). Accordingly, a mental state or process is a simulation if it resembles another state or process in some important respect and thus fulfills one of its functions (Goldman, 2009). In addition to resemblance, simulation requires the reuse of the simulator’s own mental process (Hurley, 2008). These two aspects are related, as reuse implies resemblance, provided identical input information. Together, resemblance of process and reuse of structure implies similarity of outcome. Thus, in the mirroring account, if the output of thought processes differs between self and other, this cannot be explained by simulation. Unlike mirroring, imagination, or perspective taking, does not guarantee successful replication as it consists in an attempt to replicate the target’s mental state, which can be inadequately executed (Goldman & Sebanz, 2005).

A strong case can be made for the existence of low-level simulation (Gallese, 2001; Gallese & Goldman, 1998; Gallese & Sinigaglia, 2011; Goldman, 2006, 2009; Goldman & Sripada, 2005). However, the case for high-level simulation is less clear (see Goldman, 2006: Chap. 7-8). Stich and Nichols (1995) argued that it would be possible to use simulation for simple predictions (e.g., whether one would open the fridge if one was thirsty), but not for more complex processes like choices (e.g., how much one would require to give back a lottery ticket that was bought for \$1). This argument was inspired by the prevailing wisdom in the 1990s that there would be various decision paradoxes that could not be predicted successfully by observers. To solve this difficulty,

it was proposed that irrational mental processes cannot be simulated (Heal, 1996), since many decision paradoxes are irrational, at least under a traditional view of rationality. In recent years evidence has accumulated indicating that prediction of decisions is possible in many cases (Kühberger, Kogler, Hug, & Mösl, 2006; Perner & Kühberger, 2005). However, there is little evidence that irrational decisions can be predicted successfully, and there might indeed be a profound difference between simulating rational as opposed to irrational choices. Thus, decision making, most notably in the form of decision paradoxes, maintains its appeal for distinguishing between simulation and theory.

2. Predicting Decisions

Heal (1996) proposed that simulation is possible only of the rational mind and that it is impossible to correctly predict irrational effects by using simulation. Thus, if a mental starting point (e.g., being thirsty) and an outcome (e.g., opening the fridge) are not logically related, there is no reason to assume that by simulating this process, the same starting point would lead to the same outcome. At the same time, however, Heal (1996) underlines that rationality should not be interpreted in a narrow sense depending on logical accuracy. Therefore, there could be effects that are irrational, but that are still accessible to simulation. Still, “the clearer cut the rationality, the more likely is simulation to be able to handle matters on its own” (Heal, 1996: p. 58).

This leaves us in a somewhat difficult position: some irrational effects can be simulated, others cannot. Heal gives examples of effects that she considers irrational: the position effect (Wilson & Nisbett, 1978), the Langer effect (Langer, 1975), or the endowment effect (Thaler, 1980). Although these effects are perplexing, recent research has shown that either the target effect is not replicable (Langer effect, see Kühberger, Perner, Schulte, & Leingruber, 1995), or can actually be correctly predicted. For instance, Kühberger, Kogler, Hug, & Mösl (2006) showed that the position effect (that the rightmost item is preferred when choosing from a series of identical items) becomes even stronger when participants are warned about an unknown influence of an item’s position before predicting a target’s choice. This indicates that people would in all likelihood fail in their prediction if they based it on theory, since the theory is wrong. Thus, at least some of those effects can be correctly predicted. In addition, there is evidence that this prediction is based on simulation rather than on theory.

Let us investigate a more abnormal case: delusional inference. Could delusional inference be predicted by simulation? Take, as an example, the Capgras delusion. People who suffer from the Capgras delusion take highly familiar faces to be strangers who look similar. According to a prominent cognitive neuropsychological theory (Coltheart, Langdon, & McKay, 2011; Coltheart, Menzies, & Sutton, 2010) this delusion stems from a disconnection between the face recognition system and the autonomic nervous system. This disconnection results in a discrepancy between visual and autonomic familiarity, which is solved inadequately in the—otherwise intact—belief revision system. If this view is correct, delusional inferences cannot be predicted by simulation, since the mental system of the delusional person is different from the system of the healthy person, who does not show the disconnection between face recognition and autonomic system. Delusional inference therefore might qualify as irrational, but the source is a neuroanatomical condition (disconnection of two systems). The decision-making system itself (i.e., belief revision based on new data) does not seem to be compromised. It should be noted that, evidently, delusional inferences can be predicted by theory, since psychiatrists use robust theories for making their predictions.

The critical case for Heal’s irrationality argument is different from the Capgras delusion, since it is the other way round: no neuroanatomical basis, but faulty inference. Decision paradoxes might constitute such a case. We propose to use the framing effect (Tversky & Kahneman, 1981) as the paradigmatic example, since the framing effect is often considered the prototypical irrational decision. The classic framing task, the Asian disease problem, confronts participants with the following situation: *Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.* Participants are then asked to choose under one of two different frames, where identical outcomes are described as if they were different. In the positively framed version, participants have to decide between a sure gain (Program A: 200 people will be saved) and a risky gain (Program B: a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved). In the negatively framed version, there is a sure loss (Program C: 400 people will die) and a risky loss (Program D: 1/3 probability that nobody will die and a 2/3 probability that 600 people will die).

Formally, if 200 people are saved (out of 600), this is equivalent to 400 people dying. The typical finding is, however, that participants in the positively framed condition tend to choose the sure option, whereas participants in the negatively framed condition tend to choose the risky option. Is this irrational? A standard for rational

choice is that inferences must not violate a condition called *coherence* (Adam & Reyna, 2005; Gilovich, Griffin, & Kahneman, 2002). Coherence focuses on structural relations between judgments, based on logic (deductive coherence) or probability theory (probabilistic coherence) and implies a property called *invariance*. According to invariance, choices should be independent of superficial changes in the phrasing of the options. Thus, a change in description should not alter the outcome. Accordingly, if you prefer A over B in the A-B gain pair of the Asian disease task, you are also committed to prefer Cover D in the C-D loss pair. Any inconsistency here is a violation of invariance, which is a basic postulate of rationality (Kühberger, 2002).

The framing effect has been the most impressive example among a number of irrational effects which come under the heading of “heuristics and biases”, or “cognitive illusions” (e.g., Gilovich, Griffin, & Kahneman, 2002). This tradition has successfully—probably more successfully than warranted—exported the idea that human intuitive judgment and decision making is heavily afflicted by irrationality to being widely accepted in the social and economic sciences. For a broad public, the framing effect is astonishing and irrational, since a different phrasing of the same outcomes should not lead to a different choice. This feature of the framing effect—that it is perplexing and perceived as irrational—makes it a perfect candidate for testing Heals’ proposal about ST.

In what follows we report the results of an experiment testing whether the framing effect can be predicted. If the experiment shows that the effect cannot be predicted, simulation is in trouble. Note that the boundary conditions for the actual affect and the prediction of the effect are identical: participants learn of a hypothetical situation and have to predict a choice, either for themselves, or for some other person. This equates imaginative adequacy for own and predicted choice.

It is unlikely that lay people entertain a correct theory about framing. Our laboratory has done numerous studies on framing where we frequently asked people about their thoughts in this task. Very rarely, if at all, did they come up with something that even comes close to the presumed reason for the effect (the coding of the outcomes as gains versus losses), if framing conditions were presented between groups. Even if framing was manipulated within participants (i.e., participants had to decide in the gain as well as in the loss pair of options) the gain/loss asymmetry was seldom taken to be relevant. In addition, if informed about the framing effect people often said that this choice preference came as a big surprise. Taken together, people’s reactions indicated that a general theory about the effects of framing on choice is probably not part of our folk-psychological makeup (Perner & Kühberger, 2002). That is, lay people won’t have a relevant general theory on framing available, neither explicit nor implicit, neither precise nor vague (see McKenzie & Nelson, 2003). Successful prediction of the effect therefore is unlikely to be based on theory.

We tested whether people can correctly predict the framing effect while using self-reports for investigating prediction strategy use, as it was suggested previously that this method can be valid for reporting prediction strategies (Bazinger & Kühberger, 2012a; Dunning & Hayes, 1996). As an additional variable, we manipulated the familiarity (unknown vs. well-known) of the target person, since familiarity plays a role for the choice of prediction strategies (Bazinger & Kühberger, 2013; Karniol, 2003; Nickerson, 1999). If there is little knowledge available about the target person (as it is usually the case with a stranger), there are no other options but to use general theory or simulation, as the basis for a prediction. In contrast, if there is a lot of knowledge available about the other person, it is likely that this knowledge will be applied in a prediction. Recently, it has been shown that participants do indeed use specific information when predicting well-known people (Krienen et al., 2010; Schurz et al., 2015) and that participants simulate more when predicting unknown compared to well-known people. Thus, familiarity is related to theory use in predictable ways.

3. Method

3.1. Participants

We conducted an online study sent out to the students of the University of Salzburg, which was answered by 471 respondents (121 male). Mean age was 25.35 ($SD = 7.55$) years.

3.2. Materials and Procedure

Participants were randomly divided into two groups: In the choice group, which served as a control, participants indicated their own choice in either the positively or the negatively framed task. In the prediction group, participants predicted the choice of a well-known or an unknown target person, in either the positively or the negatively

framed task. For predicting a well-known person people were provided with the following instruction: “Well-known person means that you see that person regularly and that you have a lot of contact with that person”. In the unknown person condition the instruction read: “Unknown person means that you have never met that person and that you never had contact with that person”. These definitions of degrees of familiarity are adapted from Johnson (1987). That is, familiarity was manipulated by degree of contact, rather than by social role (e.g., friend, mother, father, etc.). Thus, we used a 3 (target: self vs. unknown other vs. well-known other) \times 2 (positive vs. negative framing) between-participants design. Below is the example of the positively framed prediction task:

Some person who is unknown to you (unknown means that you have never met that person and that you never had contact with that person) is confronted with the following situation: *Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed:*

Program A: 200 people will be saved.

Program B: There is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved.

What do you predict: Which program will this person choose?

After making the predictions, participants were asked to report on their prediction strategy on a 6-point scale ranging from 0 (strongly disagree) to 5 (strongly agree). The scale was adapted from Bazinger and Kühberger (2012b) and included simulation items (“Self as Proxy”: *I chose the alternative that I personally would choose*; “Perspective Taking”: *I put myself into the other person’s position*), application of knowledge items (“General Knowledge”: *I thought about which alternative people in general would choose, and chose the alternative that I think is preferred by most people*; “Specific Knowledge”: *I have specific knowledge about the other person, and considered that knowledge in my prediction*), and an item asking whether participants had arrived at their prediction randomly (*I did not care and picked an alternative randomly*). The order of items was varied randomly.

Finally, for checking whether participants had explicit knowledge about the framing effect, we asked them whether they knew what the framing effect was (yes, no), and to provide a short description of the framing effect (in case they knew it).

4. Results

4.1. Choices and Predictions

Since this was an online study, we took care to exclude participants whose responses might be problematic. We excluded all participants who indicated having either chosen randomly, or being knowledgeable about the framing effect ($N = 26$). A sample size of 445 participants (113 male) remained, with 223 participants in the prediction group (either predicting well-known or unknown others), and 222 participants in the choice group.

The main findings are shown in Figure 1. In the choice group, we found a unidirectional framing effect: less risk-seeking in the positive than in the negative framing condition (55.6% vs. 79.7%; $\chi^2(1, N = 222) = 14.90$; $p < 0.001$; $\phi = 0.26$). For the prediction groups, similar findings were observed. When predicting an unknown person, 51% predicted the choice of the risky option in the positive framing condition, and 72.2% predicted a risky choice in the negative framing condition ($\chi^2(1, N = 112) = 4.91$; $p = 0.027$; $\phi = 0.21$). When predicting a well-known target person, 45% predicted the choice of the risky option in the positive framing condition, and 65% predicted the choice of the risky option in the negative framing condition ($\chi^2(1, N = 111) = 4.85$; $p = 0.028$; $\phi = 0.21$). To test for an interaction between target person and framing condition, we conducted a logistic regression predicting choice (sure vs. risky option, sure was coded as 0 and risky was coded as 1) from framing condition (positive vs. negative), target person (self vs. unknown vs. well-known), and the interaction between framing condition and target person. Only framing condition reached significance. Neither the target person nor the interaction between framing and target (see Table 1) reached significance. In sum, predictions in neither groups differed from own choices, and a framing effect of about the same size (notice the similar effect sizes) was evident for own choices as well as for predictions.

4.2. Reported Prediction Strategies

Figure 2 shows the means and 95% confidence intervals of reported prediction strategies. To enable easy

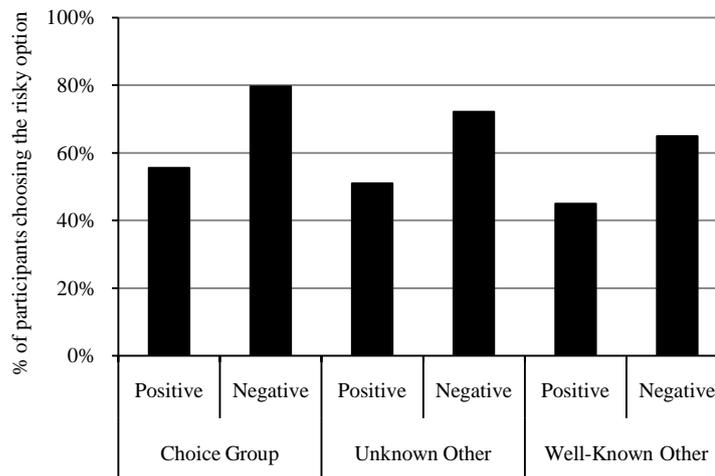


Figure 1. Percentages of participants choosing the risky option in the choice group, in the unknown target person condition and the well-known target person condition.

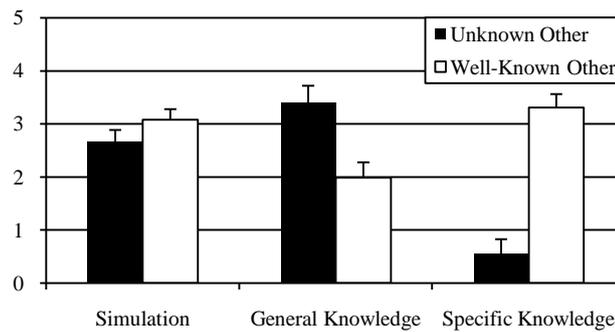


Figure 2. Means and 95% confidence intervals of reported prediction strategies for well-known and unknown target person.

Table 1. Logistic regression on choice (sure vs. risky option) as dependent variable, and framing condition, target person (self vs. unknown vs. well-known other-self was used as the baseline that the other two categories were compared to) and the interaction between framing condition and target person as independent variables.

| | B (SE) | Odds Ratio [95% CI] |
|--------------------|----------------|---------------------|
| Framing Condition | 1.19*** (0.30) | 3.29 [1.83, 5.91] |
| Unknown Other | -0.13 (0.35) | 0.87 [0.44, 1.73] |
| Well-known Other | -0.38 (0.33) | 0.69 [0.36, 1.30] |
| Unknown*Framing | -0.28 (0.51) | 0.76 [0.28, 2.07] |
| Well-known*Framing | -0.37 (0.48) | 0.69 [0.27, 1.77] |

Note: *** $p < 0.001$.

comparison, we calculated a simulation index (SI) by dividing the simulation score (i.e., the sum of the two simulation items) by the overall score (i.e., the sum of the simulation items plus the two theory items). The closer this simulation index is to 1, the more it indicates that simulation was reported as prediction strategy; the closer it comes to 0, the more it indicates that theory was reported. If SI is close to 0.5, about the same amount of simulation and theory was reported. A t-test showed a significant difference of the SI between unknown ($M = 0.43$,

$SD = 0.25$) and well-known target persons ($M = 0.38$, $SD = 0.13$); $t(219) = 2.08$, $p = 0.039$, $d = 0.26$). This shows that familiarity influenced the reported prediction strategies. Specifically, more simulation was reported for predicting unknown than well-known people. In addition, a size of SI well below 0.5 indicates that people reported more use of theory than of simulation. However, none of the strategies was reported exclusively.

4.3. Use of Theory

It is unlikely that participants predicted choices by using some exclusive theory on framing, since, presumably, no such theory is part of the folk-psychological makeup, neither explicit nor implicit. It is, however, possible that participants applied some other—however irrelevant—theory in their predictions. For instance, they could entertain the theory that people are risk-seeking in general, or that one is not allowed to gamble when lives are at stake. If a majority of predictors held such a theory, predictions would not differ between framing conditions, but would result in a preference for either the risky, or the sure option. Thus, if most people thought that other people are risk-seekers, they would predict the choice of the risky option, irrespective of framing. To test whether some general theory was used, we performed a logistic regression with prediction strategy (simulation, general knowledge, specific knowledge) and framing condition as independent variables, and predicted choice as the dependent variable. Results can be found in [Table 2](#). Only two variables reached significance: Framing condition (more risky choices were predicted in negative framing), and simulation (the less participants reported simulation, the more they predicted the choice of the sure option). General knowledge was not a significant predictor, invalidating the possibility that some general theory different from a theory on framing was responsible for the predictions. Taken together, this shows that 1) predicting by simulation resulted in predicting the majority option, that is, the option that most participants chose for themselves, and 2) people did not use some irrelevant theory consistently.

5. Discussion

We examined whether the paradoxical framing effect can be correctly predicted. In an online study we replicated the existence of the framing effect for people's own choices. In addition, the effect was also predicted for other people's choices. Self-report data on prediction strategies indicated a mix of simulation and theory.

5.1. High-Level Simulation Not Ruled Out

Our central finding is that people predicted choices of other people in a way that a framing effect was evident. Had we found the contrary, i.e., that people failed in predicting the framing effect, high-level simulation as a predictive strategy would have been ruled out. If people predict a framing effect for their own hypothetical choices, they should also predict it for other's choices based on identical hypothetical input. ST thus has survived this basic examination.

Recall that the hard-liner's position of TT holds that TT is the dominant explanatory strategy in cognitive science, and that ST is a challenge to this view ([Stich & Nichols, 1992](#)). The hard-liner's position is unlikely to be true for low-level simulation, where the evidence for simulation collected in the last 20 years or so is indeed strong (e.g., [Bach, 2011](#); [Goldman, 2006](#); [Shanton & Goldman, 2010](#)). However, high-level simulation has not yet been shown convincingly. The minimum we can say is that our findings do not rule out high-level simulation

Table 2. Logistic regression on predicted choice as dependent variable, and simulation, general knowledge, specific knowledge, and framing condition as independent variables.

| | B (SE) | Odds Ratio [95% CI] |
|--------------------|--------------------------|---------------------|
| Simulation | 0.35 [*] (0.15) | 1.41 [1.06, 1.88] |
| General Knowledge | -0.08 (0.09) | 0.93 [0.77, 1.12] |
| Specific Knowledge | -0.14 (0.09) | 0.87 [0.73, 1.03] |
| Framing Condition | 0.78 [*] (0.32) | 2.18 [1.18, 4.03] |

Note: ^{*} $p < 0.05$.

as a prediction strategy.

5.2. The Framing Paradox

From a formal viewpoint, if 600 people are at risk in total, and 200 people will be saved, this implies that 400 people will not be saved. However, there are pragmatic considerations arguing that this implication may not follow (Geurts, 2013), since people bring additional background knowledge to bear. For instance, they could interpret the sure gain option as meaning *about 200 people will be saved*, since expectations are inherently uncertain (Jou, Shanteau, & Harris, 1996; Kühberger, 1995; Macdonald, 1986). Similarly, fuzzy trace theory (Reyna & Brainerd, 1991) suggests that people do not process the numerical information in an exact way, but rather have a preference for the lowest level of processing that is necessary for solving the task at hand. In the context of framing, the quantitative information of the two options can be reduced to its qualitative gist: whether some people will be saved or not, and whether some people die or not. A preference for *some people being saved* over *some people being saved and some people not being saved* (this is the gist translation of the sure and risky gain option, respectively), and a dislike for *some people being killed* over *some people being killed and some people not being killed* (gist translation of the sure and risky loss option, respectively) leads to the framing effect. If choice in framing tasks is based on extracted gist, and if gist extraction is not treated as a distortion, the framing effect cannot be considered a bias.

Recently, information leakage was proposed as a source of the framing effect (McKenzie & Nelson, 2003; Sher & McKenzie, 2006). The basic idea is that the speaker's reference point is communicated (e.g., if people usually live or die) by his or her choice of frame. This provides the listener with choice-relevant information which can influence preferences. In the context of risky choice framing, Sher and McKenzie (2008) have argued that a speaker's choice of frame may signal that he or she considers an option to be relatively (un)successful. For instance, when people were free to frame an option, they were more likely to frame the sure option as a gain when they preferred the sure option over the risky option. Thus, their preference could be inferred from their choice of frame (see also Gigerenzer, 2015).

Pragmatic processes like gist extraction or information leakage could indeed be used for arguing that framing conditions fail to be equivalent. Similarly, many background assumptions about tacit background knowledge may also lead to the failure of equivalence (Geurts, 2013; Kühberger, 1995). Such inferences are not equally likely in different framing tasks, however. The occurrence of leakage depends on two factors (Sher & McKenzie, 2006): First, an impoverished environment—the less information there is, the more impact leakage of background information will have. Second, most scenarios provide only a fraction of a whole story, and participants have to fill in own inferences and details. Similarly, according to fuzzy trace theory, ambiguous, complex scenarios are reduced to simpler information in order to make better sense of them. A particular task that does not have these features, however, is gambling for money. In framed gambles, all relevant information is presented—outcomes and their respective probabilities—leaving no ambiguities. Still, the framing effect occurs in gambling tasks. Information leakage does not apply in a gambling setting and fuzzy trace theory probably also fails since gist extraction is unlikely for gambles, given that everybody knows that exact numerical outcomes are the essence of gambles. Thus, while recent accounts offer explanations that point to the use of pragmatic inferences in some framing tasks, the framing phenomenon as a whole cannot be easily stripped off its appearance of a paradox.

5.3. The Argument on Irrationality

High-level simulation does not guarantee correct predictions. That it is doomed to fail with irrational decisions (Heal, 1996) is unlikely, however. High-level simulation is done by running one's own decision processes offline, i.e., on pretend inputs. If one's own processes are irrational, then the result of taking the system offline will also be irrational. A simulator then will attribute irrational decisions to other people.

The literature on heuristics and biases (e.g., Kahneman & Tversky, 2000; Kahneman, Slovic, & Tversky, 1982) is a rich source of findings indicating fallacious reasoning. However, these biases are systematic. That is, choices and preferences do deviate from a rational solution, but they do so in predictable ways. There is bias, rather than randomness. The framing effect also falls under this category. It may be due to fallacious reasoning, but the majority of people follow the same fallacy. Presumably, biases can be simulated because they are what Bermúdez (2001) calls weakly normative: 1) their normal functioning is specifiable in behavioral terms (e.g., in

risky choice, risk-attitude should be independent of framing); 2) their functioning does not vary enormously across individuals (e.g., in risky choice, people usually agree on which option is more or less risky); 3) they are modular, at least in part (e.g., in risky choice, if something is considered a loss, loss aversion follows mandatorily), and 4) they break down in characteristic ways (e.g., in risky choice, gain-loss framing leads to risk-aversion for gains and risk-seeking for losses). Decisional biases are systematic, lead to consistent errors, and all ordinary people are susceptible to them - they are weakly normative. In contrast to coherence, this definition of the normative relies on process considerations, and does not take into account solely the output of a decision process. Our experiment is a test of a weakly normative process in the sense of Bermúdez which can be successfully simulated. There is hope, then, that decision paradoxes, as they are reported in the heuristics and biases literature, can be simulated.

Still, it should be kept in mind that the investigation of decision paradoxes is a fruitful approach, since, by virtue of their very nature, correct theories for these paradoxes tend to be absent in lay people (sometimes also in researchers). Thus, every instance of correct prediction of a paradoxical result adds a score for ST, given that it is unlikely that people entertain a correct theory. ST also scores on simplicity: TT needs a wealth of theories for the wealth of different decision paradoxes. Not so simulation, where one process fits them all.

We found that, when participants reported mainly using simulation rather than theory, they tended to predict the choice of the majority option. This follows from the very idea of simulation: if most people choose some option, they will also predict that others prefer this option. However, there is a problem with this reasoning. The choice of the majority option could also follow from reliance on an explicit theory, or rule, like: “Predict for others what you would choose for yourself.” Thus, the simulation would be the content of a rule, which is not simulation proper, but using theory by analogical reasoning. If so, simulation as well as theory, lead to the same behavioral result: the majority choice. Behaviorally, a decision between these two possibilities may be impossible. Other techniques like fMRI may be necessary, which have already been successfully used for showing low-level simulation and mirroring. There are such experiments, for instance [Coricelli and Nagel \(2009\)](#), [Suzuki et al. \(2012\)](#), or [Decety and Grèzes \(2006\)](#) which exploit the idea of shared recruitment. Most notably, [Schurz et al. \(2015\)](#) found in an fMRI study that also high-level simulation is a universal strategy that can be applied in different contexts, irrespective of the availability of theories. However, most predictions employ both strategies concurrently.

5.4. Truly Irrational Effects

Different from the heuristics and biases tradition, we suggest that truly irrational effects are found in psychiatry. A classic example is delusional inferences. As discussed for the Capgras delusion, two factors may be relevant: the disconnection between the face recognition system and the autonomic nervous system, and the functioning of the belief revision system ([Coltheart et al., 2010, 2011](#)). If this view is correct, ordinary people will be unable to simulate delusional inferences of delusional people. We are not aware of any data on this, but intuitively we would assume that—whatever she pretended to be the case—a healthy simulator would never find herself thinking that a well-known face is actually the face of a stranger. Simulation would not work since the mental system of the delusional person is different from the system of the healthy person.

A more interesting case is the following: would it be possible for a delusional person to simulate another person with the same delusion? The answer depends on whether or not the delusional simulator can conjure up the appropriate pretend beliefs. Can a person suffering from Capgras delusion imagine seeing a familiar face, given that he or she is unable to appreciate what it means for a face to be familiar? If the delusional person’s mental system does not provide the qualia of familiarity, pretending familiarity is logically impossible. Thus, by simulation it should be impossible to predict Capgras delusion for a simulator suffering from Capgras delusion even if the simulating person has the same mental system as the other person. Only theory will be helpful here. The Capgras delusion then could constitute a case for [Heals’ \(1996\)](#) suggestion that irrationality cannot be predicted by simulation.

What about the other way round? Could a predictor successfully predict rational choices of individuals with mental illness who are more able than the hale and hearty of adhering to the rules of rationality? [Hertwig and Volz \(2013\)](#) enumerate a series of findings showing that patients with mental illness or damage to specific brain regions (most of them related to regions that are implicated in diminished impact of emotions) are more rational than ordinary people. For instance, they do not show ambiguity avoidance, are more coherent and correspondent

in their choices, or less loss aversive. That is, they do not show classic decision paradoxes, presumably due to their illness. In these interesting cases we would not expect simulation to result in correct predictions. To conclude, the link between abnormality and rationality offers many possibilities to better understand the scope and limits of theory and simulation in mind-reading.

References

- Adam, M. B., & Reyna, V. (2005). Coherence and Correspondence Criteria for Rationality: Experts' Estimation of Risks of Sexually Transmitted Infections. *Journal of Behavioral Decision Making*, 18, 169-186. <http://dx.doi.org/10.1002/bdm.493>
- Amodio, D. M., & Frith, C. D. (2006). Meeting of Minds: The Medial Frontal Cortex and Social Cognition. *Nature Reviews Neuroscience*, 7, 268-277. <http://dx.doi.org/10.1038/nrn1884>
- Bach, T. (2011). Structure-Mapping: Directions from Simulation to Theory. *Philosophical Psychology*, 24, 23-51. <http://dx.doi.org/10.1080/09515089.2010.533261>
- Bazinger, C., & Kühberger, A. (2012a). Is Social Projection Based on Simulation or Theory? Why New Methods Are Needed for Differentiating. *New Ideas in Psychology*, 30, 228-235. <http://dx.doi.org/10.1016/j.newideapsych.2012.01.002>
- Bazinger, C., & Kühberger, A. (2012b). Theory Use in Social Predictions. *New Ideas in Psychology*, 30, 319-321. <http://dx.doi.org/10.1016/j.newideapsych.2012.02.003>
- Bazinger, C., & Kühberger, A. (2013). Empathy Gaps and Decisions: The Case of Curiosity. In C. Mohyieddini (Ed.), *Handbook of Psychology of Emotions: Recent Theoretical Perspectives and Novel Empirical Findings* (Volume 2, pp. 189-201). London: Nova Publishers.
- Bermúdez, J. L. (2001). Normativity and Rationality in Delusional Psychiatric Disorders. *Mind & Language*, 16, 493-457. <http://dx.doi.org/10.1111/1468-0017.00179>
- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78, 67-90. <http://dx.doi.org/10.5840/jphil198178268>
- Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional Belief. *Annual Review of Psychology*, 62, 271-298. <http://dx.doi.org/10.1146/annurev.psych.121208.131622>
- Coltheart, M., Menzies, P., & Sutton, J. (2010). Abductive Inference and Delusional Belief. *Cognitive Neuropsychiatry*, 15, 261-287. <http://dx.doi.org/10.1080/13546800903439120>
- Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 9163-9168. <http://dx.doi.org/10.1073/pnas.0807721106>
- Decety, J., & Grèzes, J. (2006). The Power of Simulation: Imagining One's Own and Other's Behavior. *Brain Research*, 1079, 4-14. <http://dx.doi.org/10.1016/j.brainres.2005.12.115>
- Dunning, D., & Hayes, A. F. (1996). Evidence for Egocentric Comparison in Social Judgment. *Journal of Personality and Social Psychology*, 71, 213-229. <http://dx.doi.org/10.1037/0022-3514.71.2.213>
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Vol. xiii). Cambridge, MA: The MIT Press.
- Gallese, V. (2001). The Shared Manifold Hypothesis. From Mirror Neurons to Empathy. *Journal of Consciousness Studies*, 8, 33-50.
- Gallese, V., & Goldman, A. (1998). Mirror Neurons and the Simulation Theory of Mind-Reading. *Trends in Cognitive Sciences*, 2, 493-501. [http://dx.doi.org/10.1016/S1364-6613\(98\)01262-5](http://dx.doi.org/10.1016/S1364-6613(98)01262-5)
- Gallese, V., & Sinigaglia, C. (2011). What Is So Special about Embodied Simulation? *Trends in Cognitive Sciences*, 15, 512-519. <http://dx.doi.org/10.1016/j.tics.2011.09.003>
- Gauker, C. (2005). The Belief-Desire Law. *Facta Philosophica*, 7, 121-144.
- Geurts, B. (2013). Alternatives in Framing and Decision Making. *Mind & Language*, 28, 1-19. <http://dx.doi.org/10.1111/mila.12005>
- Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology*, 6, 361-383. <http://dx.doi.org/10.1007/s13164-015-0248-1>
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511808098>
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/0195138929.001.0001>
- Goldman, A. I. (2009). Mirroring, Simulating and Mindreading. *Mind & Language*, 24, 235-252. <http://dx.doi.org/10.1111/j.1468-0017.2008.01361.x>

- Goldman, A. I., & Sebanz, N. (2005). Simulation, Mirroring, and a Different Argument from Error. *Trends in Cognitive Sciences*, 9, 320. <http://dx.doi.org/10.1016/j.tics.2005.05.008>
- Goldman, A. I., & Sripada, C. (2005). Simulationist Models of Face-Based Emotion Recognition. *Cognition*, 94, 193-213. <http://dx.doi.org/10.1016/j.cognition.2004.01.005>
- Gopnik, A. (1993). How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality. *Behavioral and Brain Sciences*, 16, 1-14. <http://dx.doi.org/10.1017/S0140525X00028636>
- Gordon, R. M. (1986). Folk Psychology as Simulation. *Mind & Language*, 1, 158-171. <http://dx.doi.org/10.1111/j.1468-0017.1986.tb00324.x>
- Heal, J. (1996). Simulation and Cognitive Penetrability. *Mind & Language*, 11, 44-67. <http://dx.doi.org/10.1111/j.1468-0017.1996.tb00028.x>
- Heal, J. (1998). Co-Cognition and Off-Line Simulation: Two Ways of Understanding the Simulation Approach. *Mind & Language*, 13, 477-498. <http://dx.doi.org/10.1111/1468-0017.00088>
- Hertwig, R., & Volz, K. G. (2013). Abnormality, Rationality, and Sanity. *Trends in Cognitive Sciences*, 17, 547-549. <http://dx.doi.org/10.1016/j.tics.2013.08.011>
- Hurley, S. (2008). Understanding Simulation. *Philosophy and Phenomenological Research*, 77, 755-774. <http://dx.doi.org/10.1111/j.1933-1592.2008.00220.x>
- Johnson, J. T. (1987). The Heart on the Sleeve and the Secret Self: Estimations of Hidden Emotion in Self and Acquaintances. *Journal of Personality*, 55, 563-582. <http://dx.doi.org/10.1111/j.1467-6494.1987.tb00452.x>
- Jou, J., Shanteau, J., & Harris, R. J. (1996). An Information Processing View of Framing Effects: The Role of Causal Schemas in Decision Making. *Memory & Cognition*, 24, 1-15. <http://dx.doi.org/10.3758/BF03197268>
- Kahneman, D., & Tversky, A. (2000). *Choices, Values, and Frames*. Cambridge: Cambridge University Press.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511809477>
- Karniol, R. (2003). Egocentrism versus Protocentrism: The Status of Self in Social Prediction. *Psychological Review*, 110, 564-580. <http://dx.doi.org/10.1037/0033-295X.110.3.564>
- Krienen, F. M., Tu, P.-C., & Buckner, R. L. (2010). Clan Mentality: Evidence That the Medial Prefrontal Cortex Responds to Close Others. *The Journal of Neuroscience*, 30, 13906-13915. <http://dx.doi.org/10.1523/JNEUROSCI.2180-10.2010>
- Kühberger, A. (1995). The Framing of Decisions: A New Look at Old Problems. *Organizational Behavior and Human Decision Processes*, 62, 230-240. <http://dx.doi.org/10.1006/obhd.1995.1046>
- Kühberger, A. (2002). The Rationality of Risky Decisions. *Theory & Psychology*, 12, 427-452. <http://dx.doi.org/10.1177/0959354302012004293>
- Kühberger, A., Kogler, C., Hug, A., & Mösl, E. (2006). The Role of the Position Effect in Theory and Simulation. *Mind & Language*, 21, 610-625. <http://dx.doi.org/10.1111/j.1468-0017.2006.00290.x>
- Kühberger, A., Perner, J., Schulte, M., & Leingruber, R. (1995). Choice or No Choice: Is the Langer Effect Evidence against Simulation? *Mind & Language*, 10, 423-436. <http://dx.doi.org/10.1111/j.1468-0017.1995.tb00022.x>
- Langer, E. J. (1975). The Illusion of Control. *Journal of Personality and Social Psychology*, 32, 311-328. <http://dx.doi.org/10.1037/0022-3514.32.2.311>
- Macdonald, R. R. (1986). Credible Conceptions and Implausible Probabilities. *British Journal of Mathematical and Statistical Psychology*, 39, 15-27. <http://dx.doi.org/10.1111/j.2044-8317.1986.tb00842.x>
- Malle, B. F., & Hodges, S. D. (2005). *Other Minds: How Humans Bridge the Divide Between Self and Others*. New York: Guilford Press.
- McKenzie, C. R. M., & Nelson, J. D. (2003). What a Speaker's Choice of Frame Reveals: Reference Points, Frame Selection, and Framing Effects. *Psychonomic Bulletin & Review*, 10, 596-602. <http://dx.doi.org/10.3758/BF03196520>
- Nickerson, R. S. (1999). How We Know—And Sometimes Misjudge—What Others Know: Imputing One's Own Knowledge to Others. *Psychological Bulletin*, 125, 737-759. <http://dx.doi.org/10.1037/0033-2909.125.6.737>
- Perner, J., & Brandl, J. L. (2009). Simulation à la Goldman: Pretend and collapse. *Philosophical Studies*, 144, 435-446. <http://dx.doi.org/10.1007/s11098-009-9356-z>
- Perner, J., & Kühberger, A. (2002). Framing and the Theory-Simulation Controversy. Predicting People's Decisions. *Mind & Society*, 3, 65-80. <http://dx.doi.org/10.1007/BF02513148>
- Perner, J., & Kühberger, A. (2005). Mental Simulation: Royal Roads to Other Minds? In B. F. Malle, & S. D. Hodges (Eds.), *Other Minds: How Humans Bridge the Divide between Self and Others* (pp. 174-189). New York: The Guilford Press.
- Ravenscroft, I. (2003). Simulation, Collapse and Humean Motivation. *Mind & Language*, 18, 162-174. <http://dx.doi.org/10.1111/1468-0017.00219>

- Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-Trace Theory and Framing Effects in Choice: Gist Extraction, Truncation, and Conversion. *Journal of Behavioral Decision Making*, 4, 249-262. <http://dx.doi.org/10.1002/bdm.3960040403>
- Saxe, R. (2005). Against Simulation: The Argument from Error. *Trends in Cognitive Sciences*, 9, 174-179. <http://dx.doi.org/10.1016/j.tics.2005.01.012>
- Schurz, M., Kogler, C., Scherndl, T., Kronbichler, M., & Kühberger, A. (2015). Differentiating Self-Projection from Simulation during Mentalizing: Evidence from fMRI. *PLoS ONE*, 10, e0121405. <http://dx.doi.org/10.1371/journal.pone.0121405>
- Shanton, K., & Goldman, I. (2010). Simulation Theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 527-538. <http://dx.doi.org/10.1002/wcs.33>
- Sher, S., & McKenzie, C. R. M. (2006). Information Leakage from Logically Equivalent Frames. *Cognition*, 101, 467-494. <http://dx.doi.org/10.1016/j.cognition.2005.11.001>
- Sher, S., & McKenzie, C. R. M. (2008). Framing Effects and Rationality. In N. Chater, & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 79-96). Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199216093.003.0004>
- Stich, S., & Nichols, S. (1992). Folk Psychology: Simulation or Tacit Theory? *Mind & Language*, 7, 35-71. <http://dx.doi.org/10.1111/j.1468-0017.1992.tb00196.x>
- Stich, S., & Nichols, S. (1995). Second Thoughts on Simulation. In M. Davies, & T. Stone (Eds.), *Mental Simulation: Evaluations and Applications* (pp. 87-108). Oxford: Basil Blackwell.
- Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., Cheng, K., & Nakahara, H. (2012). Learning to Simulate Others' Decisions. *Neuron*, 74, 1125-1137. <http://dx.doi.org/10.1016/j.neuron.2012.04.030>
- Thaler, R. (1980). Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior & Organization*, 1, 39-60. [http://dx.doi.org/10.1016/0167-2681\(80\)90051-7](http://dx.doi.org/10.1016/0167-2681(80)90051-7)
- Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211, 453-458. <http://dx.doi.org/10.1126/science.7455683>
- Wilson, T. D., & Nisbett, R. E. (1978). The Accuracy of Verbal Reports about the Effects of Stimuli on Evaluations and Behavior. *Social Psychology*, 41, 118-131. <http://dx.doi.org/10.2307/3033572>



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing a 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>