



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Framing and Rationality
Normative Assessment and Game-Theoretic Modeling

verfasst von | submitted by
Emre Karayer

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 642

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Philosophy and Economics

Betreut von | Supervisor:

Geoffrey Castillo PhD

Acknowledgements

I would like to extend my sincere gratitude to my supervisor Geoffrey Castillo for his guidance throughout this thesis. Thank you for all your invaluable insights, for your patience, and for supporting me as a young researcher.

I would like to thank Justin Leduc for always being incredibly helpful and giving great feedback. My work would not have been possible without his inspiring courses and demonstration of how to conduct interdisciplinary research. I would also like to thank Felix Pinkert for setting up an amazing P&E program in Vienna and introducing me to the depths of philosophy of economics.

Special thanks to all my other professors for everything they have taught me and my friends in the P&E program for all the interesting discussions. I am grateful to everyone who made me feel at home in Austria.

Last but not least, many thanks to my family for their ongoing support to my education and intellectual pursuits.

CONTENTS

1. INTRODUCTION.....	1
2. PRELUDE - SETTING THE STAGE.....	3
2.1. Framing Effects - Definition and Experimental Findings.....	3
2.2. Economic Rationality.....	8
2.3. Framing Effects and Violations of Axiomatic Requirements.....	12
2.4. How to Interpret Violations of Axioms of Choice.....	14
3. TIME TO CHALLENGE THE STANDARD VIEW.....	17
3.1. Two Arguments in Favor of the Rationality of Framing Effects.....	17
3.2. Normative Theories of Framing Effects.....	21
4. MODELING RATIONAL FRAMING EFFECTS WITH GAME THEORY.....	27
4.1. Why Another Model?.....	27
4.2. Main Parameters of the Game-Theoretic Model.....	29
4.3. Rational Explanation of Framing Effects with Language-based Games.....	31
5. CONCLUSION.....	34
REFERENCES.....	36
6. APPENDIX.....	40
6.1. Abstract ENG.....	40
6.2. Abstract DE.....	41

1. INTRODUCTION

Human decisions are found to be affected by the way a decision problem is presented. In other words, our judgments are often sensitive to *frames*. For instance, there is an observed difference between ground beef described as 25% fat, and ground beef described as 75% lean when it comes to our shopping decisions (Keren, 2007). Similarly, people perceive a glass described as half-full differently than a glass described as half-empty (Sher and McKenzie, 2006). This empirical finding arguably violates a foundational tenet of standard rational choice theory called the axiom of extensionality. According to that axiom, a rational agent has preferences over different options independently of how these options are described. Thus, one cannot rationally prefer “a 75% lean ground beef” over “a 25% fat ground beef”.

While many scholars argue that framing effects are always signs of irrationality, other recent critics suggest that it is not always the case and there might be good reasons we behave the way we do, including our different responses to different frames (Gold and List, 2004; Sher and McKenzie, 2008; Bourgeois-Gironde and Giraud, 2009; Mandel, 2014; Borie and Jullien, 2020; Bermudez, 2020; Fisher, 2024). The essence of the argument is that different descriptions are not arbitrary and provide additional meaningful, decision-relevant information to people’s expectations and preferences (De Boer, 2017). Furthermore, description-dependent choice behavior is not necessarily irrational given the fact that people make intuitive *pragmatic inferences* based on how the information is presented to them (Sher and McKenzie, 2006). Pragmatic inference is the practical process of reaching conclusions on what a speaker means and the conclusions can differ from the strict logical content of the said words.

In this MA thesis, my aim is to explore various attempts at accommodating framing effects within a rational framework and provide my own account addressing the strategic nature of the mechanism that underlies framing effects. Focusing on the strategic nature of the mechanism makes one realize that decisions involving framing effects are much more akin to a game between a speaker and a listener than to an agent deciding “against nature” – which is the traditional way of putting the distinct characteristic of decision theory as compared with game theory. This is an argument that has not been explicitly developed in the literature on framing effects to the best of my knowledge. Building on the work of various economists,

philosophers, and psychologists; I hope to provide a genuine contribution by presenting the foundations of a language-based game-theoretic model that can explain some normatively justifiable forms of framing effects.

In Section 2 of the thesis, I will present some well-known experimental findings of framing effects, explore the economic conception of rationality, and discuss various interpretations of violations of axioms of choice. In Section 3, I will scrutinize our understanding of framing effects and why they can be rational along with a selection of previous accounts that consider the rationality of some framing effects. In Section 4, I will provide my conceptual justification that game theory is better suited to study framing effects than decision theory, and illustrate the main parameters of a game-theoretic model that can explain rational framing effects. In Section 5, I will summarize my findings and conclude with a brief discussion.

2. PRELUDE - SETTING THE STAGE

2.1. Framing Effects - Definition and Experimental Findings

There is an overwhelming set of empirical evidence, often labeled “*framing effects*”, that shows that human decisions are highly influenced by the way a decision problem is presented (Tversky and Kahneman, 1981; Ross and Ward, 1996; Levin et al., 1998; Sonnemann et al., 2013). The key idea is that even though the logical structure of the problem remains unchanged, differences regarding how the problem is presented can be decision-relevant. Since framing effects are essentially an experimental phenomenon, I shall start by presenting some relevant empirical findings from lab and field experiments.

Consider the most well-known experiment of *risky-choice framing* known as the “Asian disease problem” initially presented by Amos Tversky and Daniel Kahneman (1981):

Problem: 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. Assume that the exact scientific estimate of the consequences of the programs are as follows:

As the experiment employs between-subjects design, the participants are randomly divided into two groups. The first group of respondents (N=152) are given the following two options:

- *If Program A is adopted, 200 people will be saved.*
- *If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.*

The second group of respondents (N=155) are given the following two options:

- *If Program C is adopted 400 people will die.*
- *If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.*

In both treatments, the participants are then asked:

Which of the two programs would you favor?

Tversky and Kahneman (1981) found that 72% of the participants faced with the “be saved” framing chose Program A which is the less risky option of saving 200 people for sure. On the other hand, when the problem is framed in terms of the number of people that will die, 78% of the participants chose Program D which is the more risky option with a $\frac{2}{3}$ probability that 600 people will die. However, under the standard interpretation, these different presentations refer to the same outcomes and define the same decision problem as they are mathematically equivalent. Thus, Program A (Program B) and Program C (Program D) are objectively identical. The change in the language alone seems to cause a reversal in the participants’ risk attitudes, and thus, their choices.

This experimental finding has long been used to illustrate human beings’ risk-averse choice behavior when faced with a gain frame (“be saved”) and risk-seeking choice behavior when faced with a loss frame (“die”). Furthermore, Druckman (2001) attempted to measure prior risk preferences to evaluate the magnitude of each framing effect. Instead of simply taking 50% as the baseline, he presented a control group that exposes the participants to both frames, suggesting that the impacts of each frame on risk attitudes should cancel out. In addition to replicating the original experiment, Druckman presented the following two programs to a random group of participants:

- *If Program A’ is adopted, 200 people will be saved and 400 people will die.*
- *If Program B’ is adopted, there is 1/3 probability that 600 people will be saved and nobody will die, and 2/3 probability that no people will be saved and 600 people will die.*

Druckman (2001) found that 43.6% of the participants faced with the survival-mortality description chose Program A’ which is the risk-averse alternative. This result indicates that people are on average slightly risk-seeking, rather than risk-neutral, in this setup. In addition, Druckman found that 68.1% of the participants faced with the original “be saved” framing chose Program A while only 22.8% of the participants chose the mathematically equivalent Program C when the problem is framed in terms of the number of people that will die. Even

though the impact of the gain frame is found to be smaller compared with Tversky and Kahneman's experiment, neither of the gain or loss frames is found to be more influential.

It is easy to come up with various redescriptions of the outcomes identified in the original problem. For instance, one can use any of the following descriptions in their characterization of the problem as Kühberger (1995) illustrates:

- A. the number of people that will be saved,
- B. the number of people that will die,
- C. the number of people that will *not* be saved,
- D. the number of people that will *not* die,
- E. the numbers of people that will be saved and that will *not* be saved,
- F. the numbers of people that will die and that will *not* die,
- G. the numbers of people that will be saved and that will die.

Indeed, one *has to* choose one of the many alternative ways that a decision problem can be presented to the decision-maker. Moreover, the decision-maker's choice depends on how what happens is described. The magnitude of the effect and whether it replicates depend on the specific design of the experiment, yet, the effect is found to be highly reliable (Kühberger, 1998). Thus, it is important to note that the presenter's choice of description is not a trivial issue. Framing effects should not be taken as a highly specific behavioral anomaly concerned with hypothetical scenarios designed for economic experiments, but as a fundamental aspect of designing and representing various kinds of decision problems.

To provide a systematic structure of the relevant empirical findings and account for what is being framed, Tversky and Kahneman (1981) distinguish three kinds of framing:

- i) framing of outcomes,
- ii) framing of acts, and
- iii) framing of contingencies.

In this thesis, I will focus on framing of outcomes which mainly affects the decision-maker's evaluation of the consequences associated with the decision problem. The main body of experimental literature concerns the framing of outcomes – presumably because of the

consequentialist tradition in economics. Levin et al. (1998) further differentiate framing of outcomes into three categories:

- i1) risky-choice framing
- i2) attribute framing
- i3) goal framing

Attribute framing refers to another class of framing of outcomes that has been well-documented in the experimental literature. Keren (2007) shows that there is an observed difference between ground beef described as 25% fat, and ground beef described as 75% lean when it comes to our shopping decisions. Keren finds that given a choice between two butchers, whose beef is advertised as 25% fat or 75% lean, respectively, most people indicate that they would buy their meat from the latter butcher. In another experiment, Leong et al. (2017) find that “a basketball player who made 40% of his shots” is rated higher than “a basketball player who missed 60% of his shots”. Similarly, people perceive a glass described as half-full differently than a glass described as half-empty (Sher and McKenzie, 2006). Thus, the decision-maker’s choice is influenced by redescribing an attribute of the objects that constitute the opportunity set from which the choice is to be made. *Goal framing*, on the other hand, consists in a redescription of the goal by which outcomes are evaluated. For instance, one can evaluate consequences of one’s actions either with the goal of “maximizing wealth” or with the goal of “avoiding any unnecessary losses” (Grüne-Yanoff, 2016).

Other kinds of framing are also relevant and have been studied in the literature. The ultimatum game can be considered to illustrate an example of how redescription of actions affects individual preferences. In the ultimatum game, there are two players. One player receives an endowment of money. This player, the proposer, is told to share the money between herself and the second player. Then, the second player, the responder, will face a decision problem with two options. The responder can either accept the offer or reject it. If the responder accepts the proposer’s offer, the money is shared in the way offered by the proposer. If the responder rejects the offer, neither of them will receive any money. Tomasino et al. (2013) show that framing the actions of the proposer as “I give you” or “I take” in the ultimatum game, triggers different neural processes and leads to otherwise different results. They find that proposers share a greater proportion of the endowment with the responder in the first frame (“I give you”) compared with the second frame (“I take”). Note that the

objective mathematical payoff structure of the game remains unchanged. Yet, the participants' decisions are significantly altered by mere linguistic changes in the description of actions.

The experiments that are discussed so far benefit from minimal framing of relevant aspects of the problem situation. In other words, they merely provide redescriptions of risks, attributes, or actions that are relevant for decision-making. In contrast, there are some framing effects that attempt to influence the context of the decision problem. For instance, Lightner et al. (2017) provide the participants with the roles of customers and bankers in one treatment of the framed ultimatum game, and a scenario of bonus gift to be 'fairly split' in the other. The further instructions and the mathematical payoff structure of the game remains the same in both treatments. As a result, they find that framing the players as bankers and customers leads to more unequal distributions of the payoff. Similarly, Ross and Ward (1996) show that rates of cooperation in the Prisoners' dilemma, a game where individual outcomes depend on whether players cooperate or not, are dependent on whether the game is called 'the Wall Street game' or 'the community game'. These results suggest that the context of a decision problem is also relevant to a person's preferences, beliefs, and thus, choices. Moreover, framing effects can also be used to influence the context in which the decisions are made.

Some economists might argue that in market conditions, the kind of "mistakes" that are associated with framing effects will be eliminated and collective economic behavior will converge to standard equilibrium prices. Since economics is about market outcomes, and not about the psychology behind individual decision-making, framing effects should not bother economists. In contrast, Sonnemann et al. (2013) test the validity of this claim by conducting various experiments on prediction markets in which "prices reflect an aggregate judgment of the likelihood of a set of possible naturally occurring events". They explore the psychological pattern of *partition dependence*, which is a simple example of framing effects and is used to illustrate how the judged probability of a future event is affected by the way the entire event space is partitioned. For instance, when evaluating the possible total points in a basketball game, one might evaluate the events $\{[0,120[; [120, 180[; [180, +\infty[\}$ or $\{[0, 150[; [150, 210[; [210, +\infty[\}$. They find that prediction markets do not eliminate partition dependence and resulting prices are influenced by such framing effects.

Overall, framing effects refer to a well-established set of empirical findings and they are often economically relevant as discussed above. They illustrate systematic and predictable patterns of human choice behavior. Yet, how to interpret various kinds of framing effects in economics is an open question. Among many economists, framing effects are taken as proof of pervasive and systematic irrationality as description-dependent choice behavior deviates from the assumptions of the standard model. One should note that *rationality* is a very loaded term, and economists have a very specific conception of rationality. Therefore, I will first present the economists' conception of rationality and the standard theory of choice in economics before turning to the normative validity of framing effects.

2.2. Economic Rationality

Economics explains and predicts phenomena as consequences of individual actions, which are themselves explained in terms of alleged reasons (Hausman, 2003). To do that in a structured and scientific manner, economic models portray agents that appear in the models to some extent *rational*. This is the core idea behind *rational choice theory* and it is tightly connected to *methodological individualism*, especially when it comes to the individualistic theories of choice in microeconomics which in turn depict the foundations of rational choice. Standard economic models based on methodological individualism propose de-psychologized explanations and apply the method of situational analysis. Thus, economic phenomena are explained by a combination of a situational model and the rationality principle (Popper, 1994). It is important to note that the rationality principle in economics has nothing to do with 'rationality as a personal attitude'. Instead, the standard conception of economic rationality only requires consistency at the level of choice. Therefore, there is no direct reference to human psychology in standard economic explanations and the economic theory of choice (which is basically rational choice theory) is built on a set of axioms of choice.

The first question of my investigation is thus: Why is the scientific practice of economics confined to a simple theory of choice with no direct reference to the underlying psychological mechanisms? The answer lies in the notion of *revealed preference theory* which is an attempt to provide the scientific credibility of economics in the 20th century. According to the positivist doctrine, which influenced economic methodology, unobservable phenomena such as mental states do not have a place within science. Paul Samuelson argued that it is not

possible to build economic sciences on a psychological concept and that it is required to shift the burden of empirical adequacy from psychology to choice behavior. Moreover, Samuelson (1947) claimed that choices capture the whole ‘empirical meaning’ of utility theory. Thus, the underlying psychological processes are black-boxed and economics can advance on the observable foundation of choices. Together with von Neumann and Morgenstern’s axiomatization (1944), Samuelson’s conception of economic rationality shifted the focus of economic explanations from reasoning behavior to choice behavior (Wheeler, 2018).

Rational choice theory (RCT hereafter) starts from the idea that we have preferences over different options. Based on this, the standard RCT imposes various axioms on those preferences such as transitivity, completeness, and independence among others. These axioms are taken as rationality requirements and satisfy various types of consistency, depending on the exact formulation of the axioms (Weirich, 2009). Yet, the normative validity of each axiom is subject to debate among philosophers. Furthermore, a decision-maker’s choice behavior can be explained as if maximizing expected utility under risk given these axioms are satisfied. This *as if* methodology, advocated by Milton Friedman (1953) in his essay *The methodology of positive economics*, is still highly influential in economics and is complementary to both Samuelson’s ultra thin conception of economic rationality based on consistency at the level of choice; and an axiomatic characterization of choice behavior.

Having presented the main methodological outlook that underlies the modern conception of economic rationality, the natural next step of my presentation is each of the abovementioned axioms that lead the way to *representation theorems* that are at the heart of standard economic explanations. As noted above, preferences can be taken as comparative judgments concerning the available options. A *weak* preference relation is represented by $\leq$. If the decision-maker evaluates option A to be at least as preferable as option B, then it can be represented as $B \leq A$ for that decision-maker. A *strict* preference relation, on the other hand, is represented by $<$. In this case, indifference is dropped and the choice behavior of a decision-maker who evaluates option A to be more preferable than option B can be represented as $B < A$. Lastly, the indifference relation is represented by $\sim$ and it is used when two options are equally preferable to the decision-maker. Based on these relations, a rational agent’s preferences should satisfy the following set of conditions.

- **Axiom 1 - Completeness**

Completeness requires that for any option A and B from the set of available options, an agent should have a (weak) preference for one of the options.

- **Axiom 2 - Transitivity**

Transitivity requires that if an agent weakly prefers A over B, and B over C, then, the agent should also weakly prefer A over C. The idea is that a rational decision-maker should not form cyclical preferences.

The axioms of completeness and transitivity are often referred to as an ‘ordering’, and can be taken as a preliminary characterisation of rational preference over options (Steele and Stefansson, 2015). According to the revealed preference theory, in which preferences merely represent observed choice behavior, the completeness axiom is inevitably satisfied since a choice has to be made to begin with. Under different interpretations of preferences that deviate from the one associated with the revealed preference theory, the normative validity of the completeness axiom becomes subject to criticism.

The transitivity axiom will be more central in a normative assessment of framing effects. Thus, one should devote more attention to the normative validity of the transitivity axiom. The idea is that if you have intransitive preferences such that you prefer A over B, B over C, and C over A; you can be exploited in the market. Consider the following case that illustrates a violation of transitivity (which is inspired by LSE, 2019). A man is planning to go to the Philippines for a summer vacation. He has narrowed down his options for accommodation to three hotels; A, B, and C. He prefers A over B since A is the closest hotel to the beach, and the view is nicer compared to B. This makes up for the fact that A offers a very tiny room. Further, he prefers C over A since C offers a fantastic view and the size of the room is bigger compared to A. This makes up for the fact that C is so far away from the beach. Yet, he also prefers B over C since B is somewhat closer to the beach compared to C, and offers a huge room. This makes up for the fact that the view from B is bad. Thus, the man in question forms a cycle in his preferences based on pairwise comparisons. ($B < A$, $A < C$, $C < B$)

The kind of example illustrated above is not very difficult to imagine. Yet, it is a clear sign of irrationality if one takes transitivity as a core criterion of rationality. In that case, the man should be willing to trade in B plus some money for A. The same goes for, A and C; the man

should be willing to trade A for C. Yet, he should also be willing to trade C for B. At the end of such a process, he ends up with the initial option of A while spending money for each trade. Consequently, such a process can in principle take place indefinitely and an agent can be turned into a *money pump* given the agent has intransitive preferences (Gustafsson, 2010). A rational agent should not be turned into a money pump, and thus, should not have intransitive preferences.

- **Axiom 3 - Continuity**

Suppose $A \leq B \leq C$. Then there is a $p \in [0,1]$ such that:

$$\{pA, (1-p)C\} \sim B$$

- **Axiom 4 - Independence**

Suppose $A \leq B$. Then for any C, and any $p \in [0,1]$:

$$\{pA, (1-p)C\} \preceq \{pB, (1-p)C\}$$

Continuity suggests that there is no outcome that is too bad not to be considered in any possible gamble. Independence, on the other hand, implies that a rational agent's evaluation of outcomes should only take into account the aspects of the lotteries that differ. Given axioms 1-4 are satisfied, von Neumann and Morgenstern (1944) show that an agent's choice behavior can be represented as if maximizing expected utility under risk. Thus, the axioms of expected utility are the axioms of rational preference (Briggs, 2014). Based on this axiomatic approach, various representation theorems have been proposed to account for actual human decisions (Ramsey, 1926; von Neumann and Morgenstern, 1944; Savage, 1972; Bolker, 1966; Jeffrey, 1990). The key takeaway from representation theorems is that observed choice behavior can be represented as if maximizing expected utility and no direct reference is given to the underlying psychological mechanisms. Overall, most standard economic models employ a utility maximization framework to describe economic choices.

Even though the axiomatic approach has been highly influential in economic theory, the axiomatic requirements of expected utility theory have been criticized for being both too permissive and too restrictive. One might argue that expected utility theory can justify any kind of choice behavior by simply redefining the options in a way that choice behavior does not violate the specified axioms. Thus, expected utility theory can be criticized for being unfalsifiable. On the other hand, one might also argue that violations of transitivity and

independence are very common in actual human decisions. Yet, standards of rationality should not be too high to deem most observed human choices irrational. The further discussions on how to interpret expected utility theory are beyond the scope of this project, yet, it should help to be aware of such discussions and various interpretations when it comes to the fundamental elements of economic theory.

Lastly, there is an implicit axiom of RCT that is not presented above. This axiom often goes unnoticed or taken onboard without careful consideration. Moreover, this axiom will be at the center of any normative assessment of framing effects, along with transitivity. This is the axiom of extensionality, or description invariance as known by most economists – even though they do not exactly refer to the same phenomenon. Having laid out the basics of axiomatic RCT, it is time to look into the normativity of framing effects due to violations of such axioms.

2.3. Framing Effects and Violations of Axiomatic Requirements

Framing effects are usually considered as signs of irrationality from the viewpoint of standard RCT. In the fourth chapter of his book *Frame it again: new tools for rational decision-making*, Bermudez (2020) points out that two arguments are generally proposed for the irrationality of framing effects. The first argument is that the kind of choice behavior represented by a preference relation in the case of framing effects violates transitivity. Since intransitive preferences can lead an agent to total destruction (money pump argument), framing effects should also be irrational if they cause intransitive preferences. I have no argument against the normative validity of the transitivity axiom here, yet, whether or not the transitivity axiom is actually violated depends on how one distinguishes between the options. Thus, it is always possible to redescribe options in a way that framing effects come across as rational. The exact rules of individuation of choice options and what the best scientific practice would be are open to debate (Fumagalli, 2020).

The second, and more crucial, argument for the irrationality of framing effects underlined by Bermudez concerns the set-theoretic foundations of RCT. Standard economic theory cannot capture, and thus, claims irrational, many framing effects because RCT has an implicit axiom called *extensionality*. According to that axiom, a rational agent has preferences over different options independently of how these options are described. Thus, it is sometimes referred to as the principle of *description invariance* in economic discussions. Arrow (1982) presents an explicit formal definition of the extensionality axiom that is relevant for RCT:

“A fundamental element of rationality, so elementary that we hardly notice it, is, in logicians’ language, its extensionality. The chosen element depends on the opportunity set from which the choice is to be made, independently of how that set is described.”

The principle of extensionality in logic means that two logically equivalent sentences can be substituted *salva veritate*, that is without loss of truth value (Bourgeois and Giraud, 2009). Moreover, the model of belief and knowledge that is used in modern economic theory is set-theoretic in nature (Moscatti, 2012). Let’s take stock of what is meant by these two sentences. Preferences, which are the starting point of standard RCT, are defined over a set of available options. The axiom of extensionality is a direct implication of the mathematical properties of the set on which individual preferences are defined. As Moscatti (2012) states, extensionality is the first and most fundamental axiom of Zermelo-Fraenkel set theory. Based on the definition presented above by Arrow, extensionality implies that two objects that are logically equivalent can be substituted.

Consider some of the examples presented in *2.1 Framing Effects - Definition and Experimental Findings*. In the so-called Asian disease problem, the observed framing effects are considered as irrational. The reason is that saving 200 people out of 600, or letting 400 people die are logically equivalent, in other words, they have the same truth value. Similarly, having made 40% of your shots is logically equivalent to having missed 60% of your shots. Lastly, there is no mathematical difference between giving ‘x’ proportion of an endowment of money to the other player, or taking ‘1-x’ proportion of that money in the ultimatum game. Even though these different descriptions lead to different results, the standard RCT does not have the tools to capture this difference because of extensionality. Therefore, if an agent’s choice behavior is altered by how the same (by which what is implied is nothing more than

logical/mathematical equivalence here) object is presented, the agent in question violates extensionality.

The key takeaway from this part is that framing effects observed in experiments are associated with violations of extensionality. Yet, how to interpret these observed violations, and thus, what it implies for the normative assessment of everyday human decisions is a different issue. This issue will be the subject of the last part of Section 2.

2.4. How to Interpret Violations of Axioms of Choice

In the last four decades, each axiom of standard RCT has been subject to test. Thus, the field of experimental economics has presented a large set of empirical evidence that illustrates that human choice behavior observed in experiments often violates the axiomatic requirements of standard RCT. In return, economists and philosophers offered various routes on what to make of these empirical findings. Fumagalli (2020) identifies three main lines of responses to the reported violations of RCT's axiomatic requirements. Modifying Fumagalli's presentation of such responses, I will instead present four lines of responses which are relevant for the normative assessment of framing effects.

The proponents of the first line of response argue that the axiomatic requirements of standard RCT are normatively valid and descriptively adequate across a wide range of decision contexts. They suggest that various violations of transitivity, independence, or extensionality are caused by the agents' inexperience in the context of limited lab experiments. The idea is that these violations disappear over time when the agents are given time and economic incentive to learn and correct their 'mistakes'. Indeed, this explanation holds for some observed violations since the participants' choice behavior may converge to the predictions of standard economic theory in experiments with repeated rounds of the same decision problem and interactions. Yet, this explanation is not sufficient to account for the kind of framing effects where different descriptions convey additional relevant information which is i) not captured by the mathematical structure of the decision problem, and ii) remains influential even in repeated games where agents have plenty of time to familiarize themselves with the problem situation and learn. The experiment by Sonnemann et al. (2013) presented in Section

2.1 shows that even experienced and sophisticated agents are prone to subtle linguistic changes in the description of the decision problem. Therefore, one must look for a different explanation regarding some framing effects that are the focus of this thesis.

The second line of response suggests that observed violations are signs of systematic and pervasive irrationality and there is no way of bringing the normative and descriptive theories into harmony. This is the standard view among many behavioral economists and it requires accepting the normative validity of expected utility theory. Thus, the task of behavioral economists is taken to provide descriptive theories to explain actual irrational human behavior. Most influential proponents of this response are Tversky and Kahneman (1988, p. 272) who argue that:

“Because framing effects and the associated failures of invariance are ubiquitous, no adequate descriptive theory can ignore these phenomena. On the other hand, because invariance (or extensionality) is normatively indispensable, no adequate prescriptive theory should permit its violation. Consequently, the dream of constructing a theory that is acceptable both descriptively and normatively appears unrealizable.”

Kahneman and Tversky (1979), while acknowledging that a rational decision-maker ought to follow the rules of expected utility maximization based on a set of standard axioms, present *prospect theory* to explain reported violations of extensionality in the case of framing effects. According to prospect theory, human agents’ choice behavior is influenced by their evaluation with respect to a certain reference point. Thus, people do not merely consider outcomes in absolute terms. Instead, they are generally concerned with relative gains and losses. Even though prospect theory can account for certain framing effects (such as the Asian disease problem), it cannot explain some framing effects where subtle linguistic changes in the description are not captured in terms of gains and losses with respect to a reference point. More importantly, prospect theory portrays a highly pessimistic depiction of practical human reasoning and implies that people often make irrational choices.

The third and fourth lines of responses, which can be taken together in their normative assessment of human judgments, suggest that a modeler should assume that humans attempt to solve problems as effectively as they can. Even though it is not essentially about framing

effects, I think that the following statement by Elinor Ostrom (1990, p. 25) captures the essence of this attitude quite well:

“Instead of presuming that some individuals are incompetent, evil or irrational, and others are omniscient, I presume that individuals have very similar limited capabilities to reason and figure out the structure of complex environments. It is my responsibility as a scientist to ascertain what problems individuals are trying to solve and what factors help or hinder them in these efforts”.

This attitude suggests that the standards of economic rationality should not be impossible to achieve for human agents. Thus, instead of deeming a wide set of human decisions as irrational, one must question the fundamentals of our normative benchmark and the common scientific practice of explaining choice behavior. Based on this understanding, the proponents of the third response argue in favor of modifying or relinquishing certain axiomatic requirements of standard RCT in line with available empirical findings. In contrast, the proponents of the fourth response aim to accommodate violations of axiomatic requirements of standard RCT by modifying how an agent’s choice options are individuated and formally represented.

In the next section, I will present various accounts that fall within the last two lines of responses. Having presented the main experimental findings of framing effects and why they are taken as signs of irrationality due to violations of extensionality; it is now time to challenge the standard view and seriously reconsider the possible form of rationality behind some framing effects.

3. TIME TO CHALLENGE THE STANDARD VIEW

3.1. Two Arguments in Favor of the Rationality of Framing Effects

Until the end of the previous section, I merely attempted to present - in a neutral way - some experimental regularities that are categorized as framing effects. Among most economists, the axiom of extensionality (or description invariance) is taken as an implicit and indispensable requirement of RCT, and framing effects are considered descriptively pervasive, mathematically intractable, and clear signs of irrationality based on the economists' standard axiomatic conception of rationality (Borie and Jullien, 2020). The standard interpretation of framing effects as normatively unjustifiable is directly related to violations of extensionality. Yet, there is no reason to accept the normative validity of extensionality without proper investigation. Indeed, as I will argue in this section, the rationality of framing effects cannot be evaluated based on *logical equivalence* which is the main implication of the extensionality axiom. Instead, the foundational principles of standard RCT should be updated to account for human agents with imperfect knowledge and limited cognitive capacities.

Rather than giving in to the common narrative of systematically flawed human reasoning, I favor the proposal of Karl Popper (1994) that we should take the rationality principle in social sciences as “adequacy of our actions to our problem situations *as we see them*”. This approach requires holding the situational model responsible instead of the rationality principle when we fail to provide a rational explanation of human choice behavior. In other words, we would learn more and provide better explanations if we attempt to understand the viewpoint of the decision-maker and incorporate it into our models before deeming reported violations of extensionality irrational. The idea is that different descriptions are not arbitrary and provide additional meaningful, decision-relevant information to people's expectations and preferences (De Boer, 2017). I shall illustrate the possible rationality of framing effects with some examples focusing on the viewpoint of the decision-maker and the informational content of the described actions and consequences. There are two influential arguments in the literature in favor of the rationality of some (but not all) framing effects to the best of my knowledge.

The first argument requires focusing on whether human agents recognize alternative descriptions of an object to be equivalent or not. This argument is put forward by Mandel (2014) who shows through an experiment that mathematical expressions are usually not understood precisely. To illustrate this, Mandel makes use of a modified version of the Asian disease problem presented in Section 2.1. In this version, a group of participants are asked to choose based on the standard framing. As expected, framing effects are replicated. However, a second group of participants are given a description of the problem which specifies that *exactly* 200 people will live, or *exactly* 400 people will die, and so on. In the second treatment, the framing effect does not replicate.

The key idea is that people often interpret mathematical descriptions as stating *lower bounds*. In other words, one might understand the statement that “200 people (out of 600) will die” as suggesting that “*at least* 200 people (out of 600) will die”. Given the empirical observation that the framing effect disappears when the quantities are specified precisely, the explanation based on lower bound interpretation seems valid. Moreover, this explanation also applies to some other kinds of description-dependent behavior. For instance, in the case of the basketball player presented in the previous section, one might understand the description of “having made 40% of his shots” as suggesting that the player made *at least* 40% of his shots. Similarly, one might interpret ground beef described as 25% fat as suggesting that it contains *at least* 25% fat and choose based on that understanding.

From the viewpoint of the decision-maker who interprets quantities as referring to lower bounds, the scenario where at least 200 people (out of 600) will live is not mathematically equivalent to the scenario where at least 400 people (out of 600) will die. If such lower bound interpretations are normatively justifiable, then, some framing effects can be rationalized since violations of transitivity and extensionality are avoided. However, one may argue that the agent simply makes a mistake in believing that the described quantities represent lower bounds. Thus, the question is essentially about who to blame for the failure of the decision-maker to represent the decision problem in the way intended by the presenter. In this project, I shall not go further into investigating this question. Instead, I will present and favor another stronger argument for the rationality of some framing effects.

The second argument in favor of the normative justification of some framing effects requires focusing on whether and when different but mathematically equivalent descriptions actually refer to the same object. The key idea is that the informational content of the decision problem may have some aspects that are not captured by the mathematical or logical structure of the problem. Sher and McKenzie (2006) present this argument for the rationality of framing effects based on “*information leakage*”. To illustrate this, Sher and McKenzie show through an experiment that participants systematically distinguish between a glass “half-full” and a glass “half-empty”. They argue that when one attribute of the glass is highlighted instead of the other, human agents are likely to take it to suggest that that attribute is higher compared to a certain reference point. Thus, human agents systematically infer that a glass described as “half-full” used to be empty, and in contrast, a glass described as “half-empty” used to be full. This observation makes even more sense combined with the experimental finding that people are more likely to describe a glass as “half-full” rather than “half-empty” when the glass was initially empty (McKenzie and Nelson, 2003).

Consider again another more recent experiment presented in Section 2.1 that perfectly illustrates the idea of “*information leakage*”. Leong et al. (2017) find that “a basketball player who made 40% of his shots” is rated higher than “a basketball player who missed 60% of his shots”. In this case, frames chosen by the agent who presents the decision problem convey information about the *relative* performance of the player in addition to his absolute performance. Indeed, it is the effective way of describing the player in the “made” frame as opposed to the “missed” frame when we want to convey that he is a successful player. Leong et al. (2017) explicitly test this intuition and find that decision-makers correctly infer the information that is leaked through the description of the decision problem.

In this framework, the agent who presents the decision problem is called the *speaker*, and the agent who makes a decision based on the problem presented to her is called the *listener* (or simply the decision-maker). In the case of the basketball player, a speaker who takes the performance of the player to be relatively successful is found more likely to present the decision problem in the “made” frame. In return, the listeners are found likely to correctly infer this implicit information and the corresponding reference points in their judgments match the reference points of the speakers.

This argument based on information leakage also applies to various cases of risky-choice framing. For instance, in the Asian disease problem, if an option is described as “200 people will be saved”; the decision-maker can reasonably infer a certain reference point where less than 200 people would be saved. Thus, it is perfectly reasonable for that agent to choose the less risky option of saving 200 people for sure. Indeed, speakers who describe the options in the “saved” frame are found more likely to prefer the less risky option of saving 200 people for sure (Sher and McKenzie, 2008). Thus, the listeners correctly infer this implicit information about the reference point conveyed by the speakers’ choice of description. In contrast, when the decision-maker is faced with an option stating that “400 people will die”; she may reasonably infer a certain reference point where less than 400 people would die, i.e. more than 200 people would be saved. Thus, it is perfectly reasonable for the agent who infers such information to choose the more risky option.

Based on these empirical findings, I favor a relatively new argument in research on decision-making that differences in the framing of outcomes or acts might bring into play additional normatively relevant information. The takeaway message is that two descriptions being semantically identical does not entail that they are pragmatically identical as well (Grüne-Yanoff, 2016). Therefore, it can be perfectly reasonable and rational for the agents to make such *pragmatic inferences* based on how the information is presented to them (Sher and McKenzie, 2006). Pragmatic inference is the practical process of reaching conclusions on what a speaker means and the conclusions can differ from the strict logical content of the said words.

One may still argue that there is no inference that can be drawn from A which cannot be drawn from B, if A and B are logically equivalent, which corresponds to the standard principle of extensionality or description invariance. Yet, Sher and McKenzie (2008) make the crucial distinction that “inferences can be drawn from the fact that the speaker said ‘A’ which cannot be drawn from the fact that the speaker said ‘B’, even if A and B are logically equivalent”. The standard model does not take into account the fact that the speaker’s choice of description is normatively relevant. Indeed, a decision problem usually does not just appear out of nowhere. A decision-maker must receive the problem from an agent of some kind, and that agent who presents the logical content of the said problem will not merely choose a description at random. Consequently, as Sher and McKenzie (2006) state:

“There is no normative problem with logically equivalent but information non-equivalent descriptions leading to different decisions”.

This distinction between “logical equivalence” and “informational equivalence” is indeed a crucial issue in our investigation. Framing effects are taken as proof of irrationality since description-dependent choice behavior violates the axiom of extensionality. Yet, extensionality is an axiom based on logical equivalence. If one takes the above analysis to be accurate, then, extensionality cannot be a criterion of rationality. It is important to note the possibility that one can formulate a principle of description invariance, that is distinct from the logical axiom of extensionality, as will be discussed later on (Fisher, 2024).

This new approach requires a shift from logical equivalence to *informational equivalence* as the normative criterion. In a basic formulation, informational equivalence obtains when no choice-relevant inferences can be drawn from the speaker’s choice of description (Sher and McKenzie, 2006). In the next part, I shall present various attempts from philosophers and economists to do away with extensionality and present a framework that takes the informational content of the decision problems more seriously. All such attempts present ways to accommodate some framing effects within a rational framework.

3.2. Normative Theories of Framing Effects

I shall start with one of the earlier attempts to rationalize framing effects. *Gold and List (2004)* propose a normative theory of bounded rationality characterized as a *reason-based* account that can rationalize framing effects by modifying how agents’ choice options are individuated and formally represented. They suggest that different presentations lead to different paths in which the propositions are considered in different orders and only some properties of the options are taken into account. Moreover, these distinct decision paths lead to different choices. For instance, in the Asian disease problem, an agent may reasonably accept the normative proposition that they should not risk no lives being saved in the “saved” frame, whereas, they could reasonably accept the normative proposition that it is unacceptable to let people die for sure in the “die” frame. Even though the initial dispositions are implicitly inconsistent, the actual dispositions accepted on the chosen decision path are

explicitly consistent. This way, framing effects observed in the experiments can be comprehensible and the agents in question may in fact be rational. Yet, Grüne-Yanoff (2016) suggests that the validity of this account depends on the actual decision-making mechanism since it presents a specific reasoning process.

One may still question the normativity of framing effects if this reason-based account proposed by Gold and List is right. Indeed, framing effects can occur only if the agent's underlying initial dispositions towards the propositions in question are implicitly inconsistent. If these implicit inconsistencies are sometimes normatively justifiable, then, description-dependent choice behavior and reported violations of extensionality cannot be criticized as being irrational. However, if implicit inconsistencies themselves are always taken as violating rationality standards, then, their framework based on *path-dependence* would not provide a genuine rationalization of framing effects. Moreover, even though path-dependence can capture certain kinds of framing effects, it does not directly account for cases where the agents make pragmatic inferences instead of merely focusing on a small subset of the described option's properties.

Next, I shall briefly present and elaborate on two attempts at explicitly fleshing out the axiom of extensionality and doing away with it as a rationality criterion. First, *Bourgeois-Gironde and Giraud (2009)* propose a formal model to bypass extensionality in the Bolker–Jeffrey decision model. They do so by incorporating a new set of information conveyed by the frame (“good news” versus “bad news”) into the decision-maker's set of beliefs concerning the decision problem. For instance, in the Asian disease problem or the case of the basketball player, different descriptions lead the decision-maker to attribute different probabilities to various background conditions that are relevant for making sense of the decision problem. Thus, Bourgeois-Gironde and Giraud present a representation theorem where extensionality is left out and logical equivalence is replaced by informational equivalence, which is a revealed notion in their account. However, their account does not present an explicit criterion to distinguish between normatively justifiable and unjustifiable violations of invariance. In other words, they present a new framework that shows that some framing effects can be rational, yet, the exact method or principle to distinguish between rational and irrational framing effects is not provided. Overall, it is important to note that their representation theorem provides a highly influential benchmark for normative assessment of framing effects

that are interpreted through pragmatic inferences made by the decision-maker as they present one of the first formal models inspired by the results of Sher and McKenzie.

In another attempt to do away with extensionality, *Bermudez (2020)* presents a framework that can rationalize framing effects based on quasi-cyclical preferences. Bermudez argues that the psychological mechanisms underlying framing effects are essential for making rational decisions when human agents are faced with sophisticated problems. To illustrate this, Bermudez gives an example from a Greek tragedy where Agamemnon has to choose between sacrificing his own daughter Iphigenia to the goddess Artemis, which would lead them to defeat the Trojans, and not sacrificing Iphigenia, which would lead them to lose the battle. According to Bermudez, Agamemnon can frame and mentally represent the first action either as *following Artemis's will* or as *murdering his daughter*. Agamemnon knows that these two descriptions refer to the same outcome. Yet, on the one hand, he favors following Artemis's will to failing his ships and people, and on the other hand, he prefers failing his ships and people to murdering his daughter. Thus, Agamemnon develops quasi-cyclical preferences.

According to Bermudez, there is no normative problem with such quasi-cyclical preferences. Thus, he argues against the axiom of transitivity as a rationality criterion. His main argument is that looking at a decision problem from different perspectives is not only helpful but also required to reach rational decisions when faced with difficult multifaceted problems (Jullien, 2022). He suggests that standard RCT misses this point as classical examples of framing effects usually focus on rather simple decision problems. Bermudez aims to establish the rationality of framing effects by defining the set on which agents have preferences as *framed outcomes*, not outcomes themselves.

Even though Bermudez presents brand-new and thought-provoking ways to think about framing effects, his argument in favor of the rationality of quasi-cyclical preferences is not fully convincing. In the above example, Agamemnon *knows* that different descriptions refer to the same object, and in principle, he is capable of conceptualizing outcomes in relevantly frame-neutral ways and making choices over these frame-neutral outcomes (Fisher, 2022). Moreover, as Guala (2022) argues, a rational decision-maker has to compare the reasons that each frame favors, and then come up with an all-things-considered evaluation of these alternative outcomes. Overall, Bermudez provides an overly permissive account of rationality that can trivialize RCT, i.e., characterize all choice behavior as rational.

The previous three accounts presented above illustrate attempts to rationalize framing effects either by i) accommodating violations of axiomatic requirements of RCT by modifying how agents' choice options are individuated and formally represented (Gold & List) or by ii) relinquishing the axiomatic requirements of extensionality and transitivity (Bourgeois-Gironde & Giraud; Bermudez). Now, it is time to consider two accounts that instead of relinquishing extensionality, modify it and provide a distinct principle of description invariance for human agents with imperfect knowledge and limited cognitive capacities.

First, *Borie and Jullien (2020)* propose *description-dependent preferences* to explain some rational framing effects. They characterize description invariance as several axioms, rather than associating it with a monolithic axiom of extensionality, and then weaken it to formalize description-dependent preferences. They define the standard axiom of description invariance as suggesting that logical equivalence of descriptions from the decision-modeller's perspective implies indifference from the rational decision-maker's perspective. Contrary to this, in their framework, the decision-maker's perspective defines equivalence classes of options. In other words, unlike the standard model where equivalence classes of options are decided based on an objective logical principle of extensionality, they define the decision-maker's preference relation on *described* outcomes.

According to the standard axiom, an agent who rates "a basketball player who made 40% of his shots" higher than "a basketball player who missed 60% of his shots" is considered irrational since these two descriptions are consequentially equivalent from the decision-modeller's perspective. Yet, if one acknowledges that the agent in question makes a pragmatic inference from the speaker's choice of description, these two descriptions are not informationally equivalent from the decision-maker's perspective. Indeed, that is what Borie and Jullien capture by proposing an axiom of description invariance that applies to described consequences from the decision-maker's perspective, not objective consequences themselves. Overall, a set of framing effects where the decision-maker makes pragmatic inferences can be rationalized since violations of transitivity and an updated version of description invariance (extensionality) are avoided.

Even though Borie and Jullien present a well-developed explicit criterion to distinguish between normatively justifiable and unjustifiable violations of invariance, their model can be criticized on a certain aspect. They argue that even if a decision-maker perceives two descriptions as being about the same object, he or she can still rationally evaluate the descriptions differently and make different choices. Yet, this approach can potentially justify any kind of choice behavior unless there are further restrictions. Instead, I suggest that the decision-maker may perceive different descriptions to be about the same object, but she needs to perceive them to be located in different *contexts* to choose differently. Different descriptions imply different informational contexts in which the speaker's judgments and relevant inferred reference points are different; and this is what justifies different evaluations. Otherwise, the observed frame-sensitive choice behavior cannot be rationalized.

Based on this understanding, I shall present the last attempt to characterize a suitably weakened axiom of description invariance that will be considered in this section. Fisher (2024) advocates for taking into account “first, agents’ epistemic situations, i.e. whether and when they recognize alternative descriptions of an object to be equivalent; and second, the individuation of objects of description, i.e. whether and when objects count as the same or different”. The first part is related to lower-bound interpretations presented by Mandel (2014) and the second part is related to the notion of “information leakage” presented by Sher and McKenzie. Thus, the main idea is to propose a shift from logical equivalence to informational equivalence as the normative criterion in a systematic way. The standard axiom of description invariance can be stated as:

- *“Any given object is evaluated in the same way under alternative descriptions.”*

On the contrary, Fisher (2024) proposes the following formulation of description invariance that can accommodate justifiable forms of description-dependence:

- *“Any given object is evaluated in the same way under alternative descriptions unless: (a) it is not known to the agent that the descriptions (as interpreted by the agent in the context of evaluation) share an object, such that (b) it is believed to be possible that the descriptions refer to distinct objects with distinct properties, and (c) those distinct properties, if they obtained, would justify different evaluations;*
or

(d) it is not known that the context of the object is the same under each of the alternative descriptions, such that (e) it is believed to be possible that there are distinct contexts with distinct properties, and (f) those distinct properties, if they obtained, would justify different evaluations.”

The conditions (d), (e), and (f) are especially relevant for the normative assessment of framing effects where the decision-maker makes pragmatic inferences about the background conditions that are relevant for making sense of the problem situation. For instance, in the case of the basketball player presented by Leong et al. (2017), the description chosen by the speaker (“having made 40% of his shots” versus “having missed 60% of his shots”) conveys information about the *relative* performance of the player, which constitutes information about the performance of other players, and thus, information about the context in which the decision problem is presented. Therefore, the decision-maker can reasonably believe that the presented frame implies a distinct context with distinct properties that would justify different evaluations. Unlike Borie and Jullien’s description-dependent preferences, different descriptions need to be distinguished from the perspective of the decision-maker to justify the rationality of such framing effects.

Among the various attempts at accommodating some framing effects within a rational framework, Sarah Fisher’s modified description invariance for human agents along with Dino Borie and Dorian Jullien’s description-dependent preferences present the most promising models if one accepts the above analysis. Yet, none of the previously proposed accounts explicitly address the distinct role of the identity of the speaker as a person with his or her own intentions, limited knowledge, and subjective preferences. Thus, I suggest that these previous attempts to enhance the normative plausibility of RCT through accommodating rational framing effects can benefit from a further refinement highlighting the strategic nature of the interaction. This will be the subject of the next section.

4. MODELING RATIONAL FRAMING EFFECTS WITH GAME THEORY

4.1. Why Another Model?

In Section 2, I have shown that framing effects are standardly taken as signs of irrationality since description-dependent choice behavior violates the axiom of extensionality. Then, in Section 3, I have presented and favored the opposing view that human agents make reasonable pragmatic inferences based on how a decision problem is presented to them. Thus, the logical principle of extensionality should be replaced by a modified principle of description invariance based on informational equivalence. Borie and Jullien (2020) and Fisher (2024) present such modifications to standard RCT. They mainly focus on the informational content of the decision problem that would justify different evaluations when agents are faced with different but logically equivalent descriptions. Yet, their models do not fully capture the role of the speaker and the kind of interaction that a listener and a speaker have. In a previous paper, Jullien (2016) underlines that what is relevant in a decision-making process is not only the communicated words but also who the speaker is (what his or her intentions are). Similarly, Fisher (2024) notes that the reliability of the speaker and the listener's corresponding beliefs are relevant for making use of implicit linguistic cues, which are set aside in her characterization of description invariance to draw the general lesson. Based on this understanding, I will offer a further refinement that would explicitly address the strategic nature of the mechanism that underlies framing effects.

Framing effects are usually modeled within a decision-theoretic framework in which agents are represented as deciding “against nature”, which is the traditional way of putting the distinct characteristic of decision theory. For instance, in the Asian disease problem, the description of the programs is simply given to a decision-maker. Similarly, in the case of the basketball player, the description of the performance of a player as “having made 40% of his shots” or “having missed 60% of his shots” is taken as the starting point of a decision-making process. Under this interpretation, there is only one decision to be made by a decision-maker based on the presented decision problem. Thus, there is no direct reference to who presents the decision problem and what their motivations are in choosing a certain way to present that decision problem.

On the contrary, a decision-maker's understanding of a problem situation is not only formed by the communicated words but also by their interaction with the speaker. After all, the speaker's choice of description is the starting point for understanding the context of a decision problem. As mentioned before, pragmatic inference is the practical process of reaching conclusions on what a speaker means and the conclusions can differ from the strict logical content of the said words. In addition, such practical conclusions also depend on the identity of the speaker.

For instance, Blount (1995) conducts an experiment with the repeated ultimatum game, and finds that human agents who have the role of responders accept lower offers from proposers if they believe that the proposer is a random number generator, rather than another human. Thus, it is obvious and quite standard in game theory to take into account a player's beliefs about the other player in explaining his or her actions. Similarly, a listener's beliefs about the speaker are relevant for assessing the value of the implicit information conveyed by the description chosen by the speaker. Thus, the *type* of the speaker (i.e. what they know of the state and what their preferences are) is also normatively relevant for a modified axiom of description invariance based on informational equivalence.

Taking this idea seriously makes one realize that decisions involving framing effects are much more akin to a game between a speaker and a listener than to an agent deciding against nature. A decision problem usually does not just appear out of nowhere. It has to be presented by an agent of some kind. Moreover, when a social agent is making a decision, he or she often considers who presented the decision problem along with the description of the said problem. The takeaway message is that the kind of interaction between the listener and the speaker is relevant for the normative assessment of framing effects.

The proposal that game theory is better suited to study framing effects than decision theory is an argument that has not been developed in the literature on framing effects to the best of my knowledge. Even if the above analysis is correct, one may still question the value of a shift from a decision-theoretic framework to a game-theoretic framework, given that the latter would complicate the model. Thus, in addition to the conceptual justification that game theory can be used to account for the strategic nature of the mechanism that underlies framing effects, practical justification of the proposed shift is required concerning the new framework's explanatory power. In the next part, I shall illustrate the explanatory power of

the proposed game-theoretic model by introducing the main ingredients of the new model that are not fully captured in the previous accounts.

4.2. Main Parameters of the Game-Theoretic Model

The modified versions of the axiom of description invariance presented by Borie and Jullien (2020) and Fisher (2024) successfully highlight the cases where description-dependent choice behavior can in principle be rational. Yet, their versions of the axiom do not filter out cases where i) agents simply make a mistake in believing that the different descriptions refer to distinct contexts with distinct properties, or ii) agents are manipulated by the speaker through choice of framing. I suggest that a game-theoretic model with two players, the speaker and the listener (decision-maker), can represent the conditions that are required for rational and justifiable forms of description-dependent choice behavior. Signaling games, that is, communication games played by two players, could be a good starting point to conceptualize this approach. For instance, the speaker can be modeled as the “*sender*” of the information, while the decision-maker can be modeled as the “*receiver*” who evaluates both the described consequences and the reliability of the speaker. The potential parameters of the new framework can be summarized as follows:

- i) the receiver’s level of knowledge in the topic of the decision problem
- ii) the receiver’s beliefs about the sender (regarding the sender’s intention and level of knowledge concerning the problem situation)
- iii) the sender’s beliefs about the receiver’s beliefs about the sender and so on.

Once again, consider the case of the basketball player where the performance of the player can either be described as “having made 40% of his shots” or “having missed 60% of his shots” by the sender. The first question is whether or not the receiver is an agent susceptible to framing effects. If not, the standard framework and axioms of choice would apply. In this project, I focus on human agents with imperfect knowledge and limited cognitive capacities who are shown to be sensitive to subtle linguistic changes in the description of a problem situation. Thus, as shown by Leong et al. (2017), human agents are likely to correctly infer the implicit information that is “leaked” by the sender. Yet, how a rational human agent ought

to assess the value of that implicit information is a separate question. There are multiple factors that affect the evaluation of the receiver.

The first factor that is relevant for the receiver's evaluation of the basketball player is the receiver's level of knowledge in the topic of the decision problem. Indeed, if the receiver is an expert on basketball and has a good understanding of what counts as a good ratio of making shots, then that receiver is less prone to framing effects. In that case, even if a rational decision-maker correctly infers the sender's judgment of the basketball player, he or she may not be influenced by the sender's choice of description.

The second factor that is relevant for assessing the value of the inferred implicit information is the receiver's beliefs about the sender's intention and level of knowledge. For instance, if the sender is known to be unreliable or manipulative, then a rational receiver's evaluation ought not to be influenced by the sender's choice of description. Moreover, some people who are found not to be influenced by the sender's choice of description may actually have formed their judgments based on the belief that the sender is trying to manipulate them. Indeed, whenever we are presented with a problem situation that is framed by a speaker, we evaluate both the described consequences and the reliability of the speaker. Such a game-theoretic model that explicitly addresses that strategic interaction between a sender and a receiver can capture these intuitions.

The third factor requires incorporating the beliefs of the sender into the model since the sender is modeled as a player, even though the sender's only action is to choose a way how to communicate the logical content of a decision problem. As mentioned, a sender does not just choose a description at random. Instead, he or she frames the problem according to his or her preferences and beliefs. Consequently, the essence of this interaction can be represented by investigating the higher-order beliefs of the two players.

The proposed game-theoretic model offers a refinement to the previous rational explanations of framing effects that imply a shift from the objective mathematical viewpoint as the normative benchmark to the subjective viewpoint of the decision-maker based on the informational content of the decision problem. Moreover, since the decision-maker must receive the decision problem from a speaker of some kind, we take no description as simply given in nature. Considering that some sort of interaction (maybe unintentional and unaware)

is inevitable, a paradigm changing implication of my approach would be that our classical problems in decision theory can be better represented in a game-theoretic framework. Yet, it is far from obvious how to formalize and test such a framework. The next part will focus on a recent approach that can provide the methodological basis for this attempt.

4.3. Rational Explanation of Framing Effects with Language-based Games

The proposed game-theoretic framework, that can account for the strategic nature of the mechanism that underlies framing effects, is part of a more general critique of outcome-based preferences. In most economic models (both standard and behavioral), preferences are defined over a set of monetary payoffs. This is tightly connected to the axiom of extensionality which implies that a rational decision-maker is indifferent between consequentially equivalent outcomes. Yet, based on a set of empirical findings that include framing effects, Capraro et al. (2024) state that “an agent’s utility depends not just on what happens, but also on how what happens is described”. Indeed, this is the main idea behind description-dependent choice behavior. Based on this understanding, Capraro et al. (2024) discuss *language-based games* where the utility function directly depends on the language. In their framework, an agent’s language describes all the features of the decision problem that are relevant to the agent. However, Capraro et al. do not present an explicit argument as to why some description-dependent behavior can be rational. I suggest that language-based games are compatible with a modified axiom of description invariance based on informational equivalence that would capture the rationality of some framing effects. Furthermore, language-based games can provide a way to formalize the proposed game-theoretic framework.

To illustrate how language-based games work, consider an example where Bob is considering proposing to his girlfriend Alice (Bjorndahl et al., 2013). Yet, Bob wants to propose only if it will be a surprise. Thus, he prefers to propose if Alice is not expecting it; and he prefers to postpone the proposal if Alice is expecting it. Capraro et al. summarize this scenario by the payoffs for Bob as shown in Table 1.

	p	$\neg p$
BA p	0	1
$\neg$ BA p	1	0

Table 1: The Surprise Proposal

Bob’s two strategies, proposing and not proposing, are denoted by p and $\neg p$, respectively, and Alice’s expecting and not expecting a proposal are denoted by BA p and $\neg$ BA p , respectively. Even though Bob is the only one making a move in this game, his utility depends on Alice’s expectations in addition to his moves. Note that Alice’s expectations are defined in Bob’s language and they can be partitioned in infinitely many different ways (for instance in 5 categories of expectation: most likely, likely, neutral, unlikely, most unlikely). The takeaway message is that such a game-theoretic model can also be used to offer a rational explanation of framing effects. I shall illustrate this possibility with an example.

Consider a game-theoretic model that can represent the case of the basketball player discussed many times in this thesis. There are two players, the sender and the receiver. The sender must choose between two descriptions of the basketball player as “a player who made 40% of his shots” or “a player who missed 60% of his shots”. Since the receiver only faces the chosen description, he or she is not aware of the full set of possible descriptions that the sender can choose. Yet, the receiver has a normatively relevant belief about the *reliability of the sender*. Similar to the example of the surprise proposal, a matrix that captures both the receiver’s strategy set (evaluating how valuable the player would be to their basketball team), and the sender’s reliability are shown in Table 2.

	not at all valuable	moderately valuable	highly valuable	extremely valuable
Reliable sender				
Unreliable sender				

Table 2: The Basketball Player

For simplicity, various factors that constitute the interaction between the sender and the receiver are captured by the receiver’s beliefs about the sender’s reliability. The payoffs would be 0 and 1, which stands for whether the receiver makes an accurate evaluation of the

described basketball player based on his level of knowledge, his beliefs about the reliability of the sender, and the value of the implicit information leaked through the description. For instance, assume that the receiver has limited knowledge of basketball. If the basketball player is described as “having made 40% of his shots” and the sender is believed to be reliable, then the rational strategy of the receiver would be to rate the player as either “highly valuable” or “extremely valuable” given that the receiver makes a correct pragmatic inference. On the other hand, consider a scenario where a receiver considers himself extremely knowledgeable about what counts as a good shooting success percentage in basketball. Moreover, the basketball player is described as “having made 40% of his shots” and the sender is believed to be unreliable and manipulative. In that case, even if the receiver correctly infers that the sender wants to convey that the description refers to a successful player, a rational receiver can discard this implicit information and evaluate the player as “moderately valuable” if that is the receiver’s own expert judgment.

One should note that the above formalization is far from a fully developed framework. Thus, it is only presented to illustrate the essence of the idea that game theory is better suited to study framing effects and that this approach can provide a more comprehensive criterion to distinguish between normatively justifiable and unjustifiable forms of framing effects. In the new model, the states of the world are replaced by the decision-maker’s beliefs about the speaker which are described in the decision-maker’s language. Unlike previous accounts, a fully developed version of this model should capture both the language and the beliefs of the decision-maker while offering a rational explanation of some framing effects based on a suitably weakened axiom of description invariance.

5. CONCLUSION

In this thesis, I aimed to challenge the standard view that framing effects are always signs of irrationality. To highlight the interdisciplinary aspect of the project, I benefited from previous research done by economists, philosophers, and psychologists. Indeed, the attempt to establish the possible form of rationality behind some framing effects can only be achieved by tightly integrating philosophy and economics. Following the existing literature on the subject, I offered my own take on the strategic nature of the mechanism that underlies framing effects. To conclude, I will summarize the three main sections of the thesis.

In Section 2, I presented some well-known experimental findings of framing effects. I provided examples to illustrate the cases of risky-choice framing, attribute framing, and the framing of actions and contexts. The crucial idea is that framing is a fundamental aspect of designing and representing various kinds of decision problems. Then, I proceeded to explain the economists' conception of rationality based on a set of axioms of choice. This axiomatization along with the standard economic methodology of explaining choice behavior *as if* maximizing expected utility is used to black-box the underlying psychological processes. Yet, framing effects are associated with violations of the axiom of extensionality (or description invariance). According to that axiom, a rational agent has preferences over different options independently of how these options are described. Thus, standard RCT does not have the tools to represent description-dependent choice behavior and deems reported violations of extensionality irrational. However, I finish Section 2 by stating that there have been various attempts to accommodate framing effects within a rational framework by either modifying the axiomatic requirements of RCT or by modifying how an agent's choice options are individuated and formally represented.

In Section 3, I presented two arguments in favor of the rationality of framing effects. Among these arguments, I favor the one that different (but consequentially equivalent) descriptions provide implicit normatively relevant information to people's expectations and preferences. As Sher and McKenzie suggest (2008) "inferences can be drawn from the fact that the speaker said 'A' which cannot be drawn from the fact that the speaker said 'B', even if A and B are logically equivalent". This new approach requires a shift from logical equivalence to *informational equivalence* as the normative criterion. Based on this understanding, I discuss

five normative theories presented by economists and philosophers to rationalize some framing effects. Among these accounts, Borie and Jullien (2020) and Fisher (2024) provide the most promising ones by developing an axiom of description invariance based on informational equivalence, which is distinct from the logical axiom of extensionality. Yet, I argue that the existing reformulations of the axiom can benefit from further refinement.

In Section 4, I presented my conceptual justification that game theory is better suited to explain rational framing effects than decision theory – an argument that has not been developed in the literature on framing effects to the best of my knowledge. Indeed, rational explanations of framing effects based on *pragmatic inferences* highlight the idea that a decision problem must be presented by an agent of some kind. Taking this idea seriously makes one realize that decisions involving framing effects are much more akin to a game between a speaker and a listener than to an agent deciding “against nature” – which is the traditional way of putting the distinct characteristic of decision theory as compared with game theory.

Moreover, I argue that the kind of interaction between a speaker, who can be modeled as the “sender” of the information”, and a listener, who can be modeled as the “receiver” of the information, is relevant for the normative assessment of framing effects. More specifically, the receiver’s beliefs about the sender’s reliability are relevant for making sense of the information that is “leaked” through the chosen description. I suggest that a language-based game-theoretic framework can account for both the language and the beliefs of the decision-maker while offering a rational explanation of some framing effects. Thus, I present the foundations of such a game-theoretic model and illustrate my approach with the case of a basketball player who can be described as “having made 40% of his shots” or “having missed 60% of his shots”.

Overall, this paper is inspired by the view that some violations of extensionality can be rational and that this needs to be accounted for through new models and experimental evidence. Even though a fully developed model is not presented in this thesis, a progressive step is taken in that direction. Further research prospects are thrilling!

REFERENCES

- Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Enquiry*, 20, 1–9.
- Bermudez, J. L. (2020). *Frame it again: New tools for rational decision-making*. Cambridge University Press.
- Bjorndahl, A., Halpern, J. Y., & Pass, R. (2013). Language-based games. *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge*, 39–48.
- Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational behavior and human decision processes*, 63(2), 131-144.
- Briggs, R. A. (2014). Normative Theories of Rational Choice: Expected Utility. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2023 ed.). Stanford University. <https://plato.stanford.edu/archives/win2023/entries/rationality-normative-utility/>
- Bolker, E. (1966). Functions Resembling Quotients of Measures. *Transactions of the American Mathematical Society*, 124(2), 292–312.
- Borie, D., & Jullien, D. (2020). Description-dependent preferences. *Journal of Economic Psychology*, 81, 102311.
- Bourgeois-Gironde, S., & Giraud, R. (2009). Framing effects as violations of extensionality. *Theory and Decision*, 67, 385-404.
- Capraro, V., Halpern, J. & Perc, M. (2024). From outcome-based to language-based preferences. *Journal of Economic Literature*, 62(1), 115-154.
- de Boer, J. (2017). Social preferences and context sensitivity. *Games*, 8(4), 43.
- Druckman, J. N. (2001). Evaluating framing effects. *Journal of economic psychology*, 22(1), 91-101.
- Fisher, S. A. (2022). Frame It Again: New Tools for Rational Decision-Making. *The Philosophical Quarterly*, 72(2), 512-514.
- Fisher, S. (2024). Description invariance: a rational principle for human agents. *Economics and Philosophy*, 40(1), 42-54.
- Friedman, M. (1953). The Methodology of Positive Economics. *Essays in Positive Economics*, 3-43.
- Fumagalli, R. (2020). On the individuation of choice options. *Philosophy of the social sciences*, 50(4), 338-365.
- Grüne-Yanoff, T. (2016). Framing. *The argumentative turn in policy analysis: Reasoning about uncertainty*, 189-215.

- Gold, N., & List, C. (2004). Framing as path-dependence. *Economics and Philosophy*, 20(2), 253–277.
- Guala, F. (2022). Consistent preferences, conflicting reasons, and rational evaluations. *Behavioral and Brain Sciences*, 25;45:e229.
- Gustafsson, J. E. (2010). A Money-Pump for Acyclic Intransitive Preferences. *Dialectica*, 64(2), 251-257.
- Hausman, D. (2003). Philosophy of Economics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2021 ed.). Stanford University. <https://plato.stanford.edu/archives/win2021/entries/economics/>
- Jeffrey, R. C. (1990). *The logic of decision*. University of Chicago press.
- Jullien, D. (2016). All frames created equal are not identical: On the structure of Kahneman and Tversky's framing effects. *Æconomia. History, Methodology, Philosophy*, (6-2), 265-291.
- Jullien, D. (2022). José Luis Bermúdez, Frame it Again: New Tools for Rational Decision-Making. *Æconomia. History, Methodology, Philosophy*, (12-1), 143-149.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Keren, G. (2007). Framing, intentions, and trust–choice incompatibility. *Organizational Behavior and Human Decision Processes*, 103(2), 238-255.
- Kühberger, A. (1995). The framing of decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes*, 62(2), 230-240.
- Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational behavior and human decision processes*, 75(1), 23-55.
- Leong, L. M., McKenzie, C. R., Sher, S., & Müller-Trede, J. (2017). The role of inference in attribute framing effects. *Journal of Behavioral Decision Making*, 30(5), 1147-1156.
- Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational behavior and human decision processes*, 76(2), 149-188.
- Lightner, A. D., Barclay, P., & Hagen, E. H. (2017). Radical framing effects in the ultimatum game: the impact of explicit culturally transmitted frames on economic decision-making. *Royal Society open science*, 4(12), 170543.
- LSE. (2019, October 15). Johanna Thoma discusses: "Money Pump," argument & rational choice | LSE Philosophy [Video]. YouTube. https://www.youtube.com/watch?v=hJ-LB0N_Aq8&t=277s

- Mandel, D. R. (2014). Do framing effects reveal irrational choice?. *Journal of Experimental Psychology: General*, 143(3), 1185.
- 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 and Review*, 10, 596–602.
- Moscatti, I. (2012). Intension, extension, and the model of belief and knowledge in economics. *Erasmus Journal for Philosophy and Economics*, 5(2), 1-26.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.
- Popper, K. R. (1994). Models, instruments, and truth: The status of the rationality principle in the social sciences. *The myth of the framework: In defence of science and rationality*, 154-184.
- Ramsey, F. P. (1926). Truth and probability. In *Readings in Formal Epistemology: Sourcebook* (pp. 21-45). Cham: Springer International Publishing.
- Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and knowledge*, 103, 135.
- Samuelson, P. (1947). *The Foundations of Economic Analysis*. Harvard University Press.
- Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.
- Sher, S., & McKenzie, C. R. (2006). Information leakage from logically equivalent frames. *Cognition*, 101(3), 467-494.
- Sher, S., & McKenzie, C. R. (2008). Framing effects and rationality. *The probabilistic mind: Prospects for Bayesian cognitive science*, 79-96.
- Steele, K., & Stefansson, H. O. (2015). Decision Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 ed.). Stanford University. <https://plato.stanford.edu/archives/win2020/entries/decision-theory/>
- Sonnemann, U., Camerer, C. F., Fox, C. R., & Langer, T. (2013). How psychological framing affects economic market prices in the lab and field. *Proceedings of the National academy of Sciences*, 110(29), 11779-11784.
- Tomasino, B., Lotto, L., Sarlo, M., Civali, C., Rumiati, R., & Rumiati, R. I. (2013). Framing the ultimatum game: the contribution of simulation. *Frontiers in human neuroscience*, 7, 337.
- Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.

- Tversky, A., & Kahneman, D. (1988). Rational choice and the framing of decisions. *Decision making: Descriptive, normative, and prescriptive interactions*, 167-192.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Weirich, P. (2009). Economic Rationality. *The Oxford Handbook of Rationality*, 380-398.
- Wheeler, G. (2018). Bounded Rationality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 ed.). Stanford University.
<https://plato.stanford.edu/archives/fall2020/entries/bounded-rationality/>

6. APPENDIX

6.1. Abstract ENG

There is an overwhelming set of empirical evidence, often labeled “framing effects”, that shows that human agents violate the axiom of extensionality (or description invariance). According to that axiom, a rational agent has preferences over different options independently of how these options are described. The standard view among behavioral economists is that framing effects are signs of systematic and pervasive irrationality of human beings. In my thesis, I challenge this standard view and argue that the rationality of framing effects cannot be evaluated based on logical equivalence which is the main implication of the axiom of extensionality. I favor a relatively new argument that different descriptions provide implicit normatively relevant information to people’s expectations and preferences. This new approach requires a shift from logical equivalence to informational equivalence as the normative criterion. Thus, I discuss some recent attempts from economists and philosophers to formulate a principle of description invariance based on informational equivalence. In addition, I suggest that a decision-maker’s understanding of a problem situation is not only formed by the communicated words but also by their interaction with the speaker. Based on this understanding, I argue that decisions involving framing effects are much more akin to a game between a speaker and a listener than to an agent deciding “against nature”. As the main genuine contribution of the thesis, I propose the foundations of a game-theoretic framework to account for rational framing effects.

6.2. Abstract DE

Es gibt eine überwältigende Anzahl empirischer Belege, die oft als “Framing-Effekte” bezeichnet werden und zeigen, dass menschliche Akteure das Axiom der Extensionalität (oder Beschreibungsinvarianz) verletzen. Diesem Axiom zufolge hat ein rationaler Akteur Präferenzen für verschiedene Optionen, unabhängig davon, wie diese Optionen beschrieben werden. Die Standardmeinung der Verhaltensökonomien ist, dass Framing-Effekte Anzeichen für eine systematische und allgegenwärtige Irrationalität des Menschen sind. In meiner Dissertation stelle ich diese Standardmeinung in Frage und argumentiere, dass die Rationalität von Framing-Effekten nicht auf der Grundlage der logischen Äquivalenz beurteilt werden kann, die die wichtigste Implikation des Axioms der Extensionalität ist. Ich vertrete die relativ neue These, dass unterschiedliche Beschreibungen implizit normativ relevante Informationen zu den Erwartungen und Präferenzen der Menschen liefern. Dieser neue Ansatz erfordert eine Verlagerung von der logischen Äquivalenz zur informationellen Äquivalenz als normatives Kriterium. Daher diskutiere ich einige neuere Versuche von Ökonomen und Philosophen, ein Prinzip der Beschreibungsinvarianz zu formulieren, das auf der informationellen Äquivalenz beruht. Darüber hinaus schlage ich vor, dass das Verständnis eines Entscheidungsträgers für eine Problemsituation nicht nur durch die kommunizierten Worte, sondern auch durch die Interaktion mit dem Sprecher geprägt wird. Auf der Grundlage dieses Verständnisses argumentiere ich, dass Entscheidungen, die Framing-Effekte beinhalten, viel eher einem Spiel zwischen einem Sprecher und einem Zuhörer ähneln als einem Agenten, der „gegen die Natur“ entscheidet. Als Hauptbeitrag dieser Arbeit schlage ich die Grundlagen eines spieltheoretischen Rahmens vor, um rationale Framing-Effekte zu erklären.

