universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Rationality as Relevance Realization:
From Axiomatic, Ecological, and Computational Rationality
to the Frame Problem of Large Worlds"

verfasst von / submitted by

Anna Riedl BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 013
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Joint Degree Programme MEi :CogSci Cognitive Science
Betreut von / Supervisor:	Mag. Dr. Johannes Jäger

Statements

Declaration of Originality

I declare that this thesis is my own work except where there is clear acknowledgment or reference to the work of others. This work has not been submitted for any other degree or professional qualification.

Statements of conjoint work

I confirm that Sections 5.5 to 6.2 were jointly co-authored with John Vervaeke, University of Toronto. I contributed 80% of this work.

Acknowledgments

I would like to thank the *Rationality Vienna* group, for years of discussing topics related to this thesis with enjoyment and intellectual virtue; Markus Peschl and the whole *MEi:CogSci programme* in Vienna and Bratislava, for teaching me unknown unknown knowledge; Johannes Jaeger, for freeing me from desert ontologies and small worlds and for his patience — he would call it process philosophy — as a supervisor; John Vervaeke (et al.), for the most relevant insight and his wisdom; and my fiance, Jakob, for our love at every step of the way.

To my parents, Rudolf and Doris

Struggling with life is a joy, thank you.

Table of Contents

1. Introduction

Rationality is roughly defined as means-ends functionality; it means to ultimately reach one's own goals and fulfilling one's values in the world (Nozick, 1993). It is usually studied as *bounded rationality*, taking into account the internal and external constraints real cognitive system face in themselves and the world (Simon, 1957; Viale, 2021). In this sense, rationality is the ability to know and doing what the right thing to do is, given one's general limited knowledge (Russell & Wefald, 1991). It is strongly linked to strategy and successful adaptation to an environment (Anderson, 1991). Historically, visions often have been to understand and increase human domain-general rationality. Fundamentally this means understanding how humans, organisms, and agents, get — something self-evidently positive, namely — what they want. More broadly, the systematic study of judgment and decision-making under uncertainty aims to find and describe principles and mechanisms underlying intelligent behavior in brains, minds, and machines (Sternberg, 2018; Gershman et al., 2015; Russell & Norvig, 2020).

Historically two principal schools of thought have studied human rationality, following opposite theories and sparked a long-standing debate called the Great Rationality Debate. One school studies rationality with the axiomatic approach, aiming for optimality modeling; the other uses the ecological approach.

The axiomatic approach is based on consistency with predefined axioms of rationality while simultaneously being their antithesis (Tverksy & Kahneman, 1974; Kahneman et al., 1982; Gilovich et al., 2002). Consequently, it assumes and postulates that judgment and decision under uncertainty can be meaningfully abstracted as a choice under risk. The conception of risk relies on the assumption that uncertainty can be quantified.

The studies on heuristics and biases based on the axiomatic approach have uncovered a multitude of deviations of human cognition from the normative models of rationality. Moreover, these results led to the formulation of empirically-grounded theories like the dual-process and prospect theory, which proposed to describe the decision-processes of the human mind (Kahneman, 2003). Furthermore, techniques to enhance human rational choice and behavior were derived from these findings, e.g., debiasing and nudging (Thaler & Sunstein, 2008). Finally, the field of behavioral economics was established based on the theories and findings of the axiomatic approach.

The ecological approach to rationality proposes a fundamentally different way of looking at what problem cognition has to solve, empathizing the adaption value to the environment and how useful heuristics are both necessary and superior for “large worlds.” However, ecological rationality disagrees with quantified risk as a useful abstraction of the fundamental and systemic uncertainty an organism faces.

The guiding question of this thesis concerns whether the conflict between the two approaches is unifiable on a deeper level by analyzing the theoretical assumptions. This thesis demonstrates how these two positions fundamentally belong in one more robust framework. This framework is guided by the ideas of computational rationality (Gershman et al., 2015) and resource-rational analysis (Griffiths et al., 2015; Lieder & Griffiths, 2020), considering the cost of computation itself and the opportunity costs for cognitive computation in the real world to determine what is rational. Computational rationality demonstrates how many heuristics that lead to biases are indeed Bayes optimal in online decision-making and real life. By approaching the phenomenon from this perspective, the thesis reaches the limitations of the computational lens, which leads to a more embedded and embodied view of expected subjective utility (Varela et al., 1991). The argumentation, therefore, bridges the optimality goal of the axiomatic approach with the focus on embeddedness and grounding of radical uncertainty in the organism-environment system of ecological rationality (Kozyreva & Hertwig, 2019).

The rationality question is firmly related to questions of neighboring disciplines (Chater et al., 2018). Therefore the insights from formulating and asking the rationality question also can influence other fields. Examples of affected topics are equilibrium analysis in economics, the conception of efficient markets, understanding intelligence in brains or engineering intelligence in machines, and the nature of empirical observation and the scientific method more broadly (Felin & Koenderink, 2021). The affected topics broadly impact the cognitive, psychological, and economic sciences. On a more practical note, a clearer understanding of rationality by synthesizing the different fronts could support finally moving beyond taking risk-modeling literal and to a more well-calibrated understanding and taming of uncertainty (Kay & King, 2020; Hertwig et al., 2019) and, outside of academic work, to improving individual and collective human rationality in general (Stanovich et al., 2016; Hertwig et al., 2019).

1.1 Aiming to Unify Axiomatic and Ecological Rationality

From the later 20th century onwards, two paradigmatic research schools around human rationality framed their studies as being built on fundamental disagreements about the interpretation of empirical facts and normative models (Stanovich & West, 2000; Tetlock & Mellers, 2002; Samuels et al., 2002; Sturm, 2012; Wallin, 2013; Petracca, 2015; Rich, 2016; Nickles, 2018; Gigerenzer, 2019).

The axiomatic approach demonstrated a systematic deviation between the normativity of optimal rationality and the description of actual human judgment and decision-making (Tverksy & Kahneman, 1974; Kahneman et al.., 1982; Gilovich et al., 2002). It accordingly associated itself with a "Meliorist" view of rationality, assuming a gap between humans and rationality that can potentially be narrowed with intervention (Stanovich, 2011). The ecological approach built its argument on the adaptional fit between organism and environment, centered on the robust value of heuristics (Gigerenzer, 1991; Todd et al., 2000; Raab, 2004; Smith, 2003; 2008; Gigerenzer, 2008; Gigerenzer et al., 2011; Brighton & Gigerenzer, 2009; Hertwig et al., 2019). As a result, it took on a "Panglossian" view of rationality, assuming the current level of rationality is "as good as it gets" and cannot fundamentally be improved.

The topic of rationality has been an interdisciplinary dialogue for decades and continues to be one. Disciplines asking relevant questions range from philosophy, economics, computer science, artificial intelligence and machine learning, complexity science, evolutionary biology, and applied mathematics to psychology and cognitive science.

The topic has been a debate for decades (Stanovich & West, 2000; Tetlock & Mellers, 2002; Samuels et al., 2002; Sturm, 2012; Wallin, 2013; Petracca, 2015; Rich, 2016; Nickles, 2018; Gigerenzer, 2019) and in its recent form, the rationality question is a renewed melting pot (Felin et al., 2014; Felin et al., 2017; Chater et al., 2018, Van Rooij et al., 2018; Otworowska et al., 2018; Gigerenzer, 2019; Brighton, 2018; 2020; Kay & King, 2020; Viale, 2021; Rich et al., 2021; Sharp et al., 2022).

The different disciplines can learn from each other and constrain each other's endeavors by considering the other perspectives. For example, potential dead-ends can be ruled out before their traveling was even attempted.

As the two sides study the same phenomenon, their research will be unifiable to the point where the content of their different perspectives stems from the foundational assumptions of the approaches. In the long run, as a result, it will be possible to differentiate knowledge about rationality generated by different disciplines from disagreements that are artifacts of different methods, perspectives, and assumptions (Giere, 2006; Wimsatt, 2007; Callebaut, 2012; Massimi & McCoy, 2019).

Some common ground has to be established to untangle the fronts of the Great rationality debate. Repeatedly misunderstandings stem from different definitions of what the term rationality means and a lack of understanding of what the respective other side argued precisely. Accordingly, this thesis will start by defining the term rationality in its weak and strong sense in Chapter One.

Chapter Two overviews the history of the Great Rationality Debate, including the main disagreements and differences between different sides and approaches.

In Chapters Three and Four, the two main perspectives on research regarding human rationality are summarized. Chapter Three focuses on axiomatic rationality. Chapter Four focuses on ecological rationality. Both Chapters Three and Four describe the underlying assumptions, the methods used, and the empirical findings. The two chapters lay crucial groundwork for the subsequent attempt at unification and discussion, as many disagreements in the Great Rationality Debate were precisely different definitions of the term "rational," a lack of understanding of the arguments of the other side or their empirical findings. For this historical reason, it is not advised to jump ahead to Chapter Five unless there are strong reasons to assume the reader already possesses all the foundational knowledge.

In Chapter Five, this work actively continues the Great Rationality Debate and argues that the axiomatic and ecological approaches can be unified into one more robust rationality question. The method used is a robustness analysis from scientific perspectivism. The re-conceptualizing of the rationality question makes it possible to exchange information more meaningfully between the axiomatic and ecological approaches. Finally, it enables the advancement of each tradition by integrating the work the respective other has done.

The reconceptualization in Chapter Five is made possible by introducing the idea of computational rationality (Gershman et al., 2015) and resource-rational analysis (Griffiths et al., 2015; Lieder & Griffiths, 2020). The thesis argues that at the center of rationality, under the radical uncertainty of "large worlds", lays the frame problem (Dennett, 1984; Hofstadter & Sander, 2013; Hoffman et al., 2015; Felin et al., 2017; 2021; Kozyreva & Hertwig, 2019; Blokpoel et al., 2018; Kay & King, 2020). The frame problem conceptualizes the rationality question as an insight problem that requires abductive reasoning and sense-making to transform "large worlds" into "small worlds" or ill-posed into well-posed problems (Knight, 1921; Savage, 1954; Simon, 1973; Marr, 1976;). The organism-environment system continuously solves the frame problem with the semiotic process of relevance realization (Vervaeke et al., 2012; Vervaeke and Ferraro, 2013; Macchi & Bagassi, 2019; Blokpoel et al., 2018).

Chapter Six draws further conclusions and examines open ends, resulting in new questions and research directions. Unfortunately, not all questions can be answered immediately, and the relevance of the argumentation to many intimately related topics is out of the scope of this work.

1.2 Definition and Foundational Knowledge Of Rationality

DAD: Son?
MOOKI: What Dad?
DAD: I've got some advice for you.
MOOKI: What's that Dad?
DAD: Do the right thing.
MOOKI: Do the right thing?
DAD: Yes.
MOOKI: That's it?
DAD: That's it.
MOOKI: OK.

Spike Lee: Do the Right Thing
as quoted in “Do the Right Thing, Studies in Limited Rationality" (Russell & Wefald, 1993)

In the ongoing clashes of the Rationality Debate since the 1980s, many of the apparent differences were due to a lack of clarifying how the term rationality is used. Therefore this work starts with some pointers on what the concept means and how it is operationalized.

As a term, rationality has many different definitions. The most significant distinction is between lay discourse usage of the word and an entirely different "strong notion" as a technical term in various disciplines like cognitive science, philosophy, economics, psychology, decision theory, and artificial intelligence (Stanovich et al., 2016; Pinker, 2021).

Historically, failing to define terms has led to long and unnecessary discussions of apparent disagreements. In the end, parts of the Great Rationality Debate in cognitive science were purely caused by semantic confusion instead of actual disagreements about phenomena in the world (Stanoivch & West, 2000).

In this work, the term refers to the strong sense of rationality typically used in decision theory and cognitive science. It roughly means means-end rationality, "reaching one's goals, whatever they are," and is usually approached in a "maximizing expected utility" notion, where utility means precisely fulfilling the agent's goals.

In its weak sense, rationality was introduced to Western philosophy by Aristotle (Aristotle, ca 40 BC), meaning roughly "based on reason" and categorizing all humans to be rational compared to animals, which are not.

This definition is not very useful, apart from the confusing fact that humans also are animals. Rationality as reason is both a binary account and a categorical notion of rationality. One can be only rational or not at all, and being part of the human species is good enough to be rational. For example, in the Cambridge Dictionary, the (weak and purely binary) definition is "the quality of being based on clear thought and reason, or of making decisions based on clear thought and reason ." None of this explains what "clear thought" or "reason" does and does not mean.

Rationality, in its strong sense, is a normative notion. Rationality means optimal judgment and decision-making (Chater & Oaksford, 1999). Additional to normative, it is therefore also specific. It has a rich foundation of formal normative ideals and is empirically and operationally grounded (Stanovich et al., 2016). It also allows for individual differences in rationality; one can be more or less rational.

It is helpful to think of their opposites to understand the difference between the strong and the weak notion of rationality. The opposite of the weak sense is not irrationality but arationality (De Sousa, 2007). Arationality means being completely "without thought" and "outside of the domain of reason," like Aristotle and philosophers in his line of thinking claim animals to be.

The opposite of the strong sense of rationality is irrationality, which does not mean the complete absence of thinking, but a mere violation of the normative idea of ideal rationality (Stanovich, 2009).

Rationality is a question of degree, defined by the distance of thought or behavior from optimality. Irrational here means purely less than perfectly rational. Most thought processes and actions are irrational in this way of conceptualizing rationality. Instead of asking "whether irrational," the exact amount of deviation — "how irrational?" — to an optimum is considered. As the word rationality comes with the heavy baggage of connotations from other fields, one could also replace it with "good thinking" at any point without losing much of the meaning (Baron, 2006; Hastie & Dawes, 2010).

Rationality, in its strong sense, has two conceptual parts, one is epistemic, and one is instrumental rationality (Stanovich, 2009). The subparts are differentiated conceptually, not mechanistically independent systems, as they are intertwined processes (Griffith et al., 2008).

Instrumental rationality is about reaching one's goals, whatever the goals are. It means behaving in the world so that one gets precisely what one most wants, given the resources (physical and mental) that are available (Hastie & Dawes, 2010). Instrumental rationality is about individual goal fulfillment, formalized as maximizing expected utility (Von Neumann & Morgenstern, 1944). This process can be easily put as decision-making or, even more accessible, as "what to do."

Nozick (1993) argued that instrumental or "means-end" rationality is the notion that lies in the intersection of all other theories of rationality and therefore is the most central one.

Epistemic rationality, on the other hand, concerns how well beliefs map onto the actual structure of the world because to take actions that fulfill one's goals, one needs to base those actions on beliefs that appropriately match the world. This epistemic part of rationality can be easier put as "judgment" or, even more straightforward — at least linguistically, not philosophically— "what is true."

Sometimes instrumental rationality is also referred to as "practical rationality", while epistemic rationality is called "evidential" or "theoretical rationality" (Foley, 1987; Harman, 1995; Manktelow, 2004; Over, 2004). A key point here is that epistemic rationality is, in a sense, instrumental to instrumental rationality and thereby supports Nozick's (1993) view on the centrality of "means-end" rationality.

The assumption of "epistemic rational" or "true" beliefs as a prerequisite for instrumental rationality is woven through all normative formalisms of axiomatic rationality. The upcoming argumentation of viewing the frame problem as central to the rationality question implicitly denies the validity of this assumption. Going back to explicitly discussing the resulting relationship between epistemic and instrumental rationality is beyond the scope of this work.

Three further points about rationality must be made to avoid confusion that can be too commonly observed:

First, rationality is not equivalent to human intelligence in the way it is commonly conceptualized in psychology. Rationality and intelligence will be further divided apart by an overview of the research by Stanovich, Toplak, and West (2016). They show the interplay of human intelligence and rationality to make individual differences in both concepts measurable. In short, rationality is measured by reaching one's goals in context; it requires fast reactions, deep thinking, and different mindware depending on the problem. Intelligence approximates the maximal processing capability but does not say how context-sensitively they are applied or how much a person uses these capabilities on average.

However, while rationality is not the same concept as in human intelligence, rationality is the intelligence aimed for by the field of artificial intelligence (AI): *"A rational agent is one that does the right thing — conceptually speaking [...] For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has." (Russell & Norvig, 2020)*

Accordingly, Russell and Norvig (2020) argue that "computational rationality" would have been the more precise term for the field of AI. In the past, McCarthy wanted to resist the term "computational" to accommodate Norbert Wiener's approach, which promoted analog cybernetics (Wiener, 1948). The wording "Artificial Intelligence", introduced at the Dartmouth conference, has stuck with the field ever since (McCarthy, 1956).

Second, rationality is not the opposite of emotions. Quite to the contrary, cognitive neuroscience has demonstrated how damage in the ventromedial prefrontal cortex (vmPFC), generally associated with emotions, leads to a pathological lack of decision-making capabilities in real life, without having impairments of sustained attention and conscious cognitive control (Bechara et al., 1994). The ventromedial prefrontal cortex is essential for representing reward and value-based decision-making through interactions with the ventral striatum and amygdala (Hiser & Koenigs, 2019).

Emotional signals are essential for rationality in an instrumental manner, as they constrain the combinatorial explosion of possible actions to a manageable number based on somatic markers stored from similar situations in the past (Stanovich et al., 2016). However, of course, they can hinder goal achievement or epistemic rationality. Nevertheless, emotions are relevant to determining what an agent cares about and adopts as its goal in the first place and help achieve those goals (Greindl, 2010; Vervaeke et al.; 2012).

Third, rationality is not the global rationality of the economic man or even omniscience. Instead, it has to be considered an optimum under external and internal constraints. Herbert Simon introduced the notion as "bounded rationality" (Simon, 1957). Bounded rationality considers the structure of the environment, the access to information, and the internal capacities and limitations possessed by the organism. Its introduction was the first step away from neoclassical economics to a microeconomic conception that fits cognitively limited beings like humans and every natural organism in the real world (Simon, 1955).

As a result of bounded rationality, not knowing something does not imply irrationality. However, acting against the limited knowledge one has would imply it. On the contrary, one could even say that wrongly basing research on the assumption that humans are omniscient would be irrational.

Other notions and forms of rationality — perfect, calculative, metalevel, bounded, computational rationality, and asymptotic bounded optimality — will be introduced and described when they become more relevant in Chapter Five.

2. The Great Rationality Debate and Its History

The phenomenon of rationality has historically attracted vivid discussions in and between disciplines. In this context, some speak of the Great Rationality Debate as a specific and singular event (Tetlock & Mellers, 2002; Stanovich, 2018), and others speak of the Rationality Wars in the plural (Samuels et al., 2002). It is an option to understand the debate in cognitive science as culminating in two peaks. The first peak would constitute the Great Rationality Debate in cognitive science between the Meliorist and Panglossion camps, which asked primarily how much rationality to attribute to humans (Stanovich & West, 2000; Tetlock & Mellers, 2002; Samuels et al., 2002; Sturm, 2012; Wallin, 2013). The second and most recent peak occurs between the axiomatic and ecological approaches to rationality, taking a step back to ask what even constitutes the rationality question and how to study it going forward (Felin et al., 2014; Felin et al., 2017; Chater et al., 2018, Van Rooij et al., 2018; Otworowska et al., 2018; Gigerenzer, 2019; Brighton, 2018; 2020; Kay & King, 2020; Viale, 2021; Rich et al., 2021;

Sharp et al., 2022). For the sake of simplicity, outside of Chapter 2 on the detailed history of the debate, this work will address the whole historical development as one debate that has significantly changed in content and participants over time.

The defining inquiry of the great rationality debate in the 20th century was whether humans are rational or not (Peterson & Beach, 1967). The answer started with an axiomatic "yes", by not even asking the question in the first place. Ultimately the answer turned into a "no" with the empirical advancements of the so-called heuristics and biases tradition (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1984; Gilovich et al., 2002). This tradition, in turn, was critically examined by the ecological approach to rationality. In recent decades, the central theme has shifted from asking whether humans are rational to asking the meta-question of "what even is the rationality question or problem" and how to conceptualize and empirically approach it (Felin et al., 2014; Felin et al., 2017; Chater et al., 2018, Van Rooij et al., 2018; Otworowska et al., 2018; Gigerenzer, 2019; Brighton, 2018; 2020; Kay & King, 2020; Viale, 2021; Rich et al., 2021; Sharp et al., 2022).

In the 1960s and early 1970s, the heuristics and biases research program started questioning the descriptive adequacy of ideal models of human judgment and decision-making that were taken as a self-evident truth in economics (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1984; Gilovich, Griffin, & Kahneman, 2002). The research program demonstrated that humans deviate systematically from rational choice theory and the axioms of rationality, which historically had a firm grip on assumptions about human information-processing capabilities across most disciplines. Accordingly, the first divide was between naive Panglossians and Meliorists. The naive Panglossian position equated normative accounts of human rationality with descriptions of actual judgment and decision-making; the Meliorist position claimed biases as systematic deviations from normative theories (Stanovich, 2011; Tetlock & Mellers, 2002).

Daniel Kahneman and Amos Tversky pioneered this historical turn of questioning economic assumptions, established empirical cases of human irrationality, and highlighted the incompatibility of human thinking with the idea of homo economicus. In 2002 Daniel Kahneman was awarded a Nobel Prize "[...] for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty" and for founding the field of behavioral economics (Thaler & Sunstein, 2008; Baddeley, 2017). This scientific development in its historical context and the friendship between Amos Tversky and Daniel Kahneman was documented in the book "The Undoing Project" (Lewis, 2016). Their Meliorist stance established that humans stray in predictable ways, so-called biases and fallacies, from classical formal benchmarks of rationality in judgment and choice. They pinpoint sharp ontological break points in human phenomenology that have no meaningful equivalents in formal models of rationality and propose new descriptive theories of cognition, the dual-process framework, and prospect theory, to capture these phenomena. Additionally, proposed theories are framing effects, mental accounting, and the psychophysics of gains and losses (Dawes, 1998; Kagel & Roth, 1995; Mellers et al., 1998; Sunstein, 2000, Tetlock & Mellers, 2002).

The following major shift in the debate took place in the early 1990s (Stein, 1996). The initial Panglossian position held by economists, which demanded a heavy burden of proof for deviations from rational choice theory, was now joined by adaptationist evolutionary psychology. The Panglossian position holds the stance that "human rationality is as good as it gets" (Stanovich, 2011) and assumes that, under constraints, "we live in the best – most rational – possible world". The new support for this position started with Gigerenzer (1991) arguing that cognitive illusions are only a byproduct of the experimental designs, namely by how probabilities are represented. Kahneman and Tversky (1996) again pointed out that cognitive illusions are actual phenomena to be taken seriously.

The two main ongoing controversies were around questions of empirical and normative boundary conditions (Tetlock & Lerner, 1999; Tetlock & Mellers, 2002), with vivid discussions between "traditionalists" versus "revisionists." Consequently, mathematical revisionists continued to loosen the axioms of expected utility theory or supplement expected-utility theory with alternative frameworks such as correspondence standards (Luce, 2000). Pragmatic standards instead emphasized "what works" in achieving long-term evolutionary success (Gigerenzer, 1996; Hammond, 1996) or adaption to the sociocultural environment (Tetlock, 2000; Galesic et al., 2012).

Gigerenzer's criticism pointed precisely towards this lack of assessment of overall ecological validity (Brunswik, 1955) of heuristic processes, which is about the correlation of the actual outcome of actions in the world with the information and cues available to the agent (Stein, 1996; Gilovich, Griffin & Kahneman, 2002; Gigerenzer, 1991; Cosmides, Tooby, 1995). The new "fast and frugal" position and the ideas of ecological rationality emerged. Conceptually it situated and embedded the agent as an organism in an actual environment (Gigerenzer & Todd, 1999; Cosmides & Tooby, 1995; Gigerenzer, 2000; Gigerenzer & Selten, 2001; Gigerenzer, 2008; Gigerenzer, 2010; Todd & Gigerenzer, 2012; Gigerenzer, Hertwig & Pachur; 2011; Gigerenzer, 2015), giving new fire to the debate between the Panglossian and Meliorist positions (Tetlock & Mellers, 2002; Gilovich, Griffin, & Kahneman, 2002; Samuels, Stich, & Bishop, 2002).

During the first decades of the debate about rationality, the focal point between these positions was the question of whether or not humans are to be considered rational.

In recent years of the ongoing discussion, this focal point changed. The prominent positions transformed into axiomatic rationality and optimality modeling on one side (Griffith & Tenenbaum, 2006; Lieder & Griffith, 2020), containing most of the ideas and scholars of the Meliorist stance; and ecological rationality related to the Panglossian stance on the other side (Gigerenzer, 2019; Brighton, 2018). This melting pot had its peak in 2018 (Chater et al., 2018) when researchers from diverse backgrounds – cognitive science, applied and experimental psychology, behavioral economics, and biology – published their perspectives on the argumentation by Felin, Koenderink, and Krueger (2017). In their assessment, Chater et al. (2018) offer a list of objections and complementations to their argument concerning rationality, perception, and the all-seeing eye (Felin et al., 2017). The exchange reconceptualized the rationality question and questioned the axioms of rationality. The covered issues have far-reaching implications for rationality research and the intersecting literature of the cognitive, psychological, and economic sciences more broadly (Chater et al., 2018).

The three underlying topics of the debate are, firstly, the notion of uncertainty.
It is argued that understanding uncertainty as risk and attempting to measure it is only valid in small axiomatic worlds. In the "large world" of ill-posed problems, organisms face radical uncertainty (Kozyreva & Hertwig, 2019; Kozyreva, Pleskac, Pachur & Hertwig, 2019; Kay & King, 2020). From this follows the consequential limitations of the axioms of rationality and the need for new statistical foundations. Notions of optimality and normativity lose importance (Brighton, 2019; Gigerenzer, 2019, Kozyreva & Hertwig, 2019; Hertwig et al., 2019), and perception acquired new importance in rationality, leaving many assumptions from psychophysics and an omniscient view "from nowhere" behind (Felin et al., 2017; Wimsatt, 2007). The new conception has consequences for the philosophy of science and the relationship between epistemic and instrumental rationality (Wimsatt, 2007; Hoffmann, Singh, & Prakash, 2015; Felin et al., 2017; Felin, 2021); these topics will get further addressed through the main argumentation in Chapter Five.

Next in this work are Chapters Three and Four. These Chapters aim to transport further foundational understanding of the axiomatic and ecological approach to rationality. They are written to avoid typical past misunderstandings of debates around the topic. In case of robust foundational knowledge of the literature the reader may continue directly to Chapter Five. The following Chapter – Chapter Three – overviews the axiomatic approach to rationality.

3. Axiomatic Rationality

Amos Tversky and Daniel Kahneman revolutionized the academic study of human judgment and decision-making in the 1960s and early 1970s with their contributions to the so-called "heuristics and biases" research program (Gilovich & Griffin, 2002). Much attention was drawn to their work when Kahneman received the Nobel Prize in 2002 "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty." (NobelMedia, 2020), and in 2011 Kahneman published the popular science book "Thinking, Fast and Slow" summarizing their work (Kahneman, 2011). In 2016 the popular science "The Undoing Project: A Friendship that Changed our minds" (Lewis, 2017) gave an even closer insight into the research and the friendship that made large fractions of this advancement possible. Similarly, the field of behavioral economics, most famous for the idea of nudging, was built on the heuristics and biases tradition and received broad public interest (Camerer et al., 2005; Thaler & Sunstein, 2008; Laibson & List, 2015; Heukelom, 2014; Heukelom & Sent, 2017). Additionally, analogous fields like behavioral law and neuroeconomics are emerging and growing.

After the replication crisis in the social sciences and psychology's central role in the second decade of the 21st century, one might conclude that many findings of the mentioned research program can be discounted retrospectively (Schooler, 2014; Diener & Biswas-Diener, 2017; Shrout & Rodgers, 2018). However, Ruggeri et al. (2020) replicated the empirical findings in one extensive meta-study, suggesting that such worries can be safely ignored.

The axiomatic approach underlying the heuristics and biases tradition must be understood with the normative theories, it is both based on and the antithesis to, in mind: The idea of humans as rational agents formalized as humans maximizing expected utility.

The research program of neoclassical economics dominated the decades following World War II (e.g., Davis, 2003; Colander et al., 2004; Sent, 2006). It was centered around the idea of the self-interested agent that maximizes the satisfaction of their preferences. The average rationality of the economically acting individual enabled the assumption that the economy as a total would be in equilibrium (Sent, 2018).

This notion of maximizing expected utility was operationalized by consistency with the axioms of rationality such as transitivity, completeness, and independence (Luce & Raiffa, 1957). Judgment and choice that follows logic, the probability calculus, the axioms of choice, and other normative theories of decision theory, meant behaving "as if" they maximized expected utility (Ramsey, 1926; Van Neumann & Morgenstern, 1944; Binmore, 2008; Al-Najjar & de Castro, 2011; Pawitan & Lee, 2021). Historically psychologists suggested, based on the shadow cast by rational choice theory (Gilovich & Griffin, 2002), that humans are "intuitive statisticians" and follow Bayesian decision theory (Peterson & Beach, 1967; Titelbaum, 2022; 2022).

The Heuristics and Biases program in the 1970s challenged this theory. The program was built on the insight that judgment under uncertainty is based on simplifying rules (heuristics), which generally are a potent and valuable strategy but also lead to systematic errors in reasoning, named biases (Kahneman & Tversky, 1974; Gilovich & Griffin, 2002). Moreover, the research program did not only question the descriptive adequacy of ideal models of judgment in economics by listing heuristics and biases as systematic deviations from normative theories. Instead, its main contribution was offering a cognitive architecture and descriptive theories as alternatives to explain the mistakes made (Gilovich & Griffin, 2002).

These insights led to neoclassical economics having to take a step back against a new pluralism of emerging economic research programs (Davis, 2007), with behavioral economics as the most prominent contender (Heukelom, 2010; Sent, 2018).

This Chapter on the axiomatic approach to rationality will be continued in three parts — a theoretical, empirical, and concluding one.

First, it will take a closer look at the theoretical underpinnings, namely the underlying normative assumptions of rational choice and decision theory and how the axiomatic approach compares human judgment to normative ideals.

The empirical part will continue to overview the empirical and theoretical research by Tversky and Kahneman and the heuristics and biases research tradition. Kahneman (2003) summarised this work with Amos Tversky, who passed away in 1996, as follows:

"My work with Tversky comprised three separate programs of research, some aspects of which were carried out with other collaborators. The first explored the heuristics that people use and the biases to which they are prone in various tasks of judgment under uncertainty, including predictions and evaluations of evidence (Kahneman and Tversky, 1973; Tversky and Kahneman, 1974; Kahneman et al., 1982). The second was concerned with prospect theory, a model of choice under risk (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) and with loss

aversion in riskless choice (Kahneman et al., 1990, 1991; Tversky and Kahneman, 1991). The third line of research dealt with framing effects and their implications for rational-agent models (Tversky and Kahneman, 1981, 1986)."

All three strains of their shared work are built on a two-system architecture of cognition, the dual-process theory.

The empirical part further includes the improvements Keith Stanovich and colleagues made to the theories of Kahneman and Tversky. He advanced the dual-process theory into the tripartite model of the mind, from which a concrete taxonomy of errors and their causes resulted. Some concluding remarks will close the Chapter as a third part.

3.1 Neoclassical Economics and the Axioms of Rationality

3.1.1 Rationality, Optimality and Rules for "Good Thinking."

After world war II and under the influence of neoclassical economics, many disciplines implicitly assumed humans to be "homo economicus" and to behave following rational choice theory.

Proponents of this view acknowledge that people make mistakes, but random ones (Gilovich & Griffin, 2002).

The model of rational choice assumes humans follow the axioms of rationality, which include, next to others, the elementary rules of logic (inductive and deductive reasoning), probability theory, Bayesian inference, and regression analysis, resulting in the claim that humans are "intuitive statisticians" or "intuitive Bayesians" (Gilovich and Griffin, 2002).

The model of economic man assumes the following three points: Decision-makers are fully informed regarding all possible options for their decisions and of all possible outcomes of their decision options, they are infinitely sensitive to the subtle distinctions among decision options, and they are entirely rational regarding their choice of options (Edwards, 1963; Tversky & Slovic, 1974; Slovic, 1991; Sternberg & Sternberg, 2012).

The quality of a decision cannot be determined by the outcome alone. Instead, it is about the potential outcomes, their probabilities, and the values to the agent at the time the choice is made. Axiomatic rationality is about the process, and the principles followed. The narrow technical implies the following four criteria have to be considered (Hastie & Dawes, 2010):

1. The choice is based on the decision maker's current assets and state.
2. It is based on the possible consequences of choice.
3. When consequences are uncertain, their likelihood is evaluated following the basic rules of probability theory.
4. The choice is adaptive within the constraints of those probabilities and the values or satisfactions - utility - associated with the possible outcomes.

If the criteria of logic and the probability calculus are not followed, the agent comes to contradictory conclusions, which can result in a so-called money pump as an analogy for losing against reality. In analytical philosophy and bayesian epistemology, this is called the Dutch book argument (Ramsey, 1926; Von Neumann & Morgenstern, 1944; Titelbaum, 2016). The philosophical argument is that propositions about reality cannot be true and false simultaneously because reality itself is not contradictory; instead, contradictory thinking is irrational (Dawes and Hastie, 2010). This assumption can and will be debated in later Chapters in the context of computationally limited agents in an uncertain and real-time environment.

3.1.2 The Axiomatic Approach in Psychology and Behavioral Economics

"Mind the gap."
London Underground, since 1968

In practice, it is difficult to assess rationality; it is experimentally challenging to measure the utilities of various consequences of actions and even more so to determine the concrete actions that are optimal ("maximizing utility") under given circumstances and preferences. The "axiomatic approach" solves this problem precisely; it offers a way to measure the rationality of decisions. This approach is called the "axiomatic approach" to whether people are maximizing utility (Stanovich, West, & Toplak, 2016). The defining criterion of the axiomatic approach is that when people's preferences follow specific rules, the axioms of rational choice, then they behave "as if" they are maximizing utility and are rational (Dawes, 1998; Edwards, 1954; Ellsberg, 1961; Jeffrey, 1983; Luce & Raiffa, 1957; Savage, 1954; Kreps, 1988; Fishburn, 1970; von Neumann & Morgenstern, 1944; Stanovich & West, 2007).

This approach allows for diagnosing deviations from the optimal choice behavior and therefore serves as an inverse measure of the degree of rationality. The more mistakes a person makes in applying the axioms of choice, the less the person is rational. The opposite applies as well. Fewer inconsistencies with the axioms of rationality mean a closer approximation to optimality and, therefore, rationality (Stanovich et al., 2016). This approach is closely related to optimality modeling and the so-called Meliorist side of the Great Rationality Debate. "Meliorist" means an implicit assumption of a substantial gap between human behavior and optimality. This stance has also led to various attempts at closing the gap with interventions like debiasing (Stanovich, 1999, 2004).

Contributors to the heuristics and biases research tradition have repeatedly been criticized for their emphasis on errors, but Meliorists argue that other domains follow analogical approaches. For example, medicine never defines "perfect health," but nevertheless offers good understanding of disease as a deviation from health and resulting practical ways to bring a human body closer to health. So do other professions. Lawyers identify injustice without defining "perfect justice" (Postman, 1988). Following the argument pointing to biases without having to specify the best behavior in any situation is a reasonable approach.

The Meliorist, axiomatic, or optimality stance on rationality in the great rationality debate conceptualizes theories in the domain of rationality in three categories (Mellers, 1990; Baron, 2008):

1. The *normative notions* of how a rational agent does or ought to reason, think and behave. The leading contender for a normative theory is rational choice theory. Earlier it was understood as a positive scientific theory, but this conception changed to a normative notion when challenged by descriptive accounts of human behavior (Hands, 2013)
2. The *descriptive notions* capture how humans judge, choose and behave. The theories usually stem from experimental psychologists.

3. The *prescriptive part* of the research tries to answer how the diagnosed gaps can be closed. Knowing the formalisms of the normative notions alone does not automatically mean behaving as if one was following them.

3.2 Empirical Advancements: From Biases to Prospect Theory

3.2.1 The Heuristics and Biases Research Program

The Heuristics and Biases Research program examined human judgment empirically and related it to its departure from a normative standard of rational choice theory.

The program examined algorithms humans use — heuristics[1] — to transform complex tasks into simpler judgemental operations and the resulting errors (Kahneman & Tversky, 1974). The three main heuristics humans use when making judgments under uncertainty are representativeness, availability, and adjustment of anchoring. Each simplifying heuristic can lead to a wide range of biases. (Kahneman & Tversky, 1974):

In everyday life, humans use the representative heuristic to assess how likely an object belongs to a particular class, how likely something originates from a specific process, and how likely a

---

[1] The term heuristic started to get widely used in computer science and the early days of artificial intelligence (Russell & Norvig, 2020). The need for heuristics in computer science and Artificial Intelligence stems from the fundamental problem of computational complexity. In computer science, the complexity of an algorithm is the amount of resources required to run it. The complexity of a problem is defined as the complexity of the simplest algorithms suitable to solving it. Calculating through all possible solutions implies high time complexity (how long it takes to find a solution) as well as space complexity (memory needed to perform the search) to guarantee completeness (finding a solution when there is one) and optimality (finding the optimal solution) of the search algorithm. Heuristics are optimally efficient when considering search cost or the total cost of an algorithm to evaluate the performance. This fundamental demonstrates the speed-accuracy trade-off and why Herbert Simon suggested satisficing instead of maximizing, as the search for the best solution is no longer the best when factoring in computation cost (Russell & Norvig, 2020).

process will generate a particular phenomenon. For example, to answer whether something belongs to a particular group, the similarity of the object to a typically imagined representation of the group is judged. Empirical data shows that people order occupations judged by probability and similarity in precisely the same way. For example, biases that can arise from using the representative heuristic are insensitivity to the prior probability of outcomes or sample size, the illusion of validity, and the misconception of regression to the mean.

To judge the number of occurrences of a class type or the probability of an event, humans often rely on the availability heuristic, which uses the sensation of ease by which the class or event comes to mind. However, in cases where the availability is not a valid cue, the heuristic leads to predictable biases. As a result, perceived risks deviate from actual risks (Slovic et al., 1981; Slovic, 1987).

Human estimates are often assessed by taking an initial value and adjusting it given other pieces of information to a final judgment. For example, the initial value may stem from the problem formulations or other unreliable sources, and the final answer is skewed towards the initial number. Experiments show the anchoring effect even when the initial number is openly generated by a randomized process like turning a fortune wheel in front of the participant's eyes.

Human decisions are based on the, at least rough, assessment of believed likelihoods of uncertain events. Heuristics reduce the complexity of these tasks to simper operations of judgment. While generally valuable, they can lead to systematic error, sometimes leading to severe deviations of optimal judgment and choice. This implies failing to reach one's goals when accepting the assumptions of the axiomatic approach to rationality (Kahneman & Tversky, 1974).

Whether heuristics are helpful depends on the validity of the data used to judge another property of the environment. For example, using the sensory input of how sharp or blurry an object appears as a cue to infer distance is usually beneficial. When the same heuristic is used under conditions that make everything appear more blurry, like fog, distance is systematically overestimated, which can, for example, lead to car accidents (Kahneman & Tversky, 1974; Kahneman et al., 1982).

The many researchers that worked on the heuristics and bias tradition identified a multitude of additional heuristics and biases. Describing more of them is, however out of the scope of this work.

3.2.2 The Dual-Process Framework

While psychology and other behavioral sciences have never been able to create as simplistic and elegant models as economists, it has offered empirically-grounded and integrative concepts that can explain a magnitude of phenomena in the real world (Kahneman, 2003). More important than identifying heuristics and biases was their explanation by understanding their cognitive underpinnings (Evans, 2017), from which different behavioral models arise. Apart from the heuristics and biases research program, other lines of research conducted by Kahneman and Tversky (Kahneman, 2003) were modeling human choice under risk with prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Kahneman et al., 1990, 1991;

Tversky and Kahneman, 1991) and framing Effects and the implications for rational-agent models (Tversky and Kahneman, 1981, 1986).

The different lines of research were all built on a dual-process computational model of cognition (2003). Therefore to fully grasp the unified treatment of judgment and choice that Kahneman and Tversky offer, the distinction between two generic modes of cognition has to be introduced, corresponding approximately to intuition and reasoning (Kahneman, 2003). Or, as the title of the popular science book suggests, "Thinking, Fast and Slow" (Kahneman, 2011).

Historically philosophical work on thinking by Aristoteles was associated only with conscious reflection, but from Plato onwards, there are many accounts of a rough understanding that thinking can take place in an unconscious, automatic or intuitive way (Evans, 2017). In their dual-process theory, Kahneman and Tversky differentiate between two similarly divided modes of thinking.

These different modes of thinking are noticeable without psychological experiments. Imagine anything you can do without effortfully thinking at every step, for example driving the car, the bicycle, or walking. In these cases, you can use your attention elsewhere, for example, to have a conversation, think about a problem from work, or listen to a podcast (Dawes & Hastie, 2010). When an unusual situation emerges, like someone cutting in line in front of you while driving, you pull your attention away from thinking about something else or talking, and you focus on solving the new traffic situation instead. The metaphor of "paying attention "conceptualizes the economic trade-off of using limited executive power to reach one's goals well.

Kahneman differentiates System 1 and System 2 as the two systems in the dual-system theory. System 1 is fast thinking, or what one would refer to as intuition. Its defining feature is autonomy. System 1 processes execute automatically upon encountering their triggering stimuli. Other correlated but non-defining features as associativeness and small load on the central processing capacity (Stanovich & Toplak, 2012). The category includes processes of emotional regulation, solving specific adaptive problems, processes of implicit learning, and automatic firing of over-learned associations (see Barnett & Kurzban, 2006; Carruthers, 2006; Evans, 2008, 2009; Morrs & de Houwer, 2006; Samuels, 2005, 2009; Shiffrin & Schneider, 1977). System 1 encompasses innately specified processing modules or procedures and experiential associations that have been learned to automaticity, like implicit learning and conditioning, as well as expert intuitions (Stanovich et al., 2016).

System 2 is the more effortful way of thinking, approximating the idea of reasoning (Kahneman, 2011). It is relatively slow and computationally expensive, so it needs attention and effort (Kahneman, 1973).

One key difference is that System 1 operates in parallel, but System 2 is largely serial.

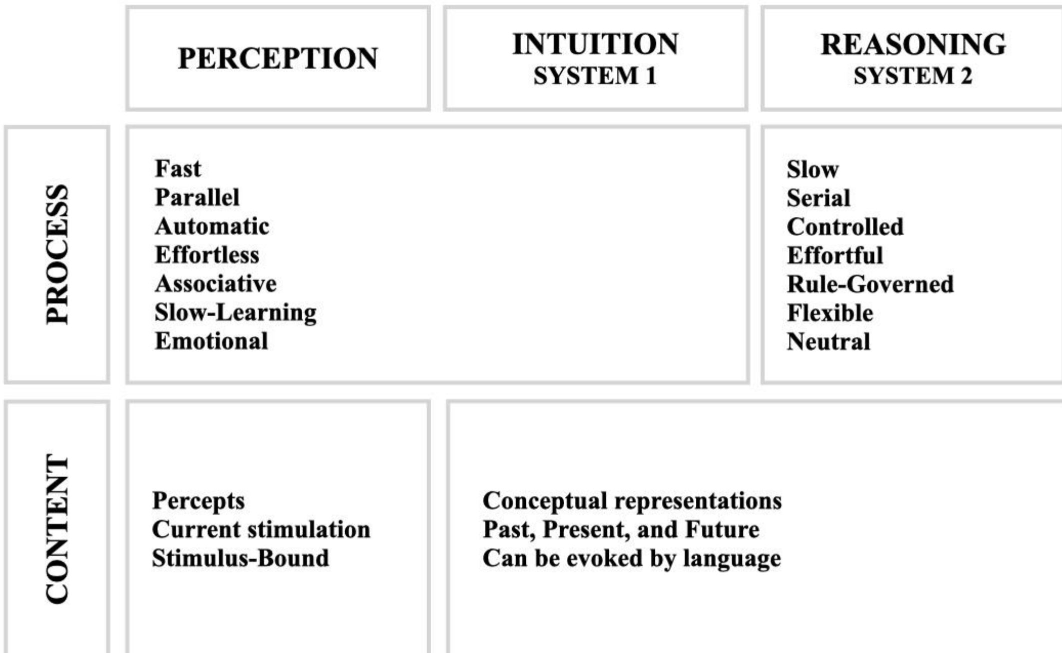

Figure 1. A depiction of the three cognitive systems — perception, System 1 and System 2 — that underpin dual-process theory. All theories empirically built on the heuristics and biases research tradition build on this architecture of cognition (Kahneman & Tversky, 1974; Kahneman et al., 1982; Kahneman, 2003).

The dual-process theory, therefore, differentiates between an associative system, which involves mental operations based on observed similarities and temporal continuities, and a rule-based system, which involves manipulations based on the relations among symbols (Barrett et al., 2004; Sloman, 1996). A detailed listing of their relative properties can be found in Figure 1. A wealth of results supports dual-process approaches since the launch of the Heuristics and Biases research program in the 1970s (Evans & Stanovich, 2013).

Talking about System 1 and System 2 (Kahneman, 2011) is not meant to be understood as a neurological claim of corresponding systems in the brain but as a cognitive architecture. Stanovich (1999) decided upon this problem and for additional clarity in the wording to speak of

Type 1 and Type 2 processes (Stanovich, 1999). While the exact properties of the two types of processing are still lively debated (De Neys et al., 2018), the overall distinction is built on a magnitude of evidence. Other scientists have used slightly different terms to refer to analogous dual-process theories (see table in Evans, 2008).

The same modes of thinking apply in less mechanical (than walking or driving a car) and more intellectual situations. Imagine the question „1 + 1 = ???“. The answer simply "pops up" or "appears" in your head. Quite to the contrary, it would probably even be difficult to not have the answer automatically "appearing "or "coming to you . "No effort whatsoever is necessary. Now think of “434 + 87 = ???” (Dawes & Hastie, 2010). You will notice that the solution will not come up automatically; you would now have to stop reading this text to use your limited attention to effortfully compute this question with a controlled, deliberate step-by-step algorithm (Dawes & Hastie, 2010)

The most important behaviors happen when it is unclear which system of speed to use and when mistakes happen in the choice of system. For example, one of the most critical functions of Type 2 processing is overruling Type 1 processing when it is mistaken. This override sometimes fails. The tests used to determine those cases are called "cognitive reflection tasks ."
Three of the original tests were introduced by Frederick (2003) and encompass the bat-and-ball-problem, the widget problem, and the lily pad problem:

- A bat and a ball cost 1.10 in total. The bat costs 1.00 more than the ball. How much does the ball cost?

- If it takes 5 machines 5 minutes to make 5 widgets. How long would it take 100 machines to make 100 widgets?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake. How long would it take for the patch to cover half of the lake?

They cause a distinct temporal pattern of processes in an average human mind: The intuitive answer for the bat and ball-problem is typically 10 cents because 1.10 divide naturally into 10 cents, and 10 cents are roughly in the ballpark of the anticipated right solution (Kahneman, 2003).

In the widget problem, an intuitive pattern-matching between "5, 5, 5" and "100, 100, X" suggests the solution is to give the same number three times and causes the reflex to respond to the question with 100.

The lily pad problem makes the number 48 the most relevant information to store in working memory and then asks about "half "of something, which intuitively suggests the answer is 24.

The correct answers are, of course, 0.05 dollars, 5 minutes, and on the 47th day.

According to Frederick, over 50 percent of his Princeton University students responded by giving the wrong answer (Frederick, 2003). This high error rate depicts how little cognitive self-monitoring of System 1 output takes place and how seldomly people think hard (Kahneman, 2003).

The operational definition of rational thinking in cognitive science leads to conceptualizing rationality within this dual-process framework. Most experiments in the heuristics and biases literature were created to pit an automatically triggered answer against a more effortfully generated (mostly, but not always) normative response (Stanovich, 2016).

Reading this, one might — intuitively! — conclude that System 2 is the better one. This conclusion is wrong; type 2 processes are not necessarily superior, as they are limited and slow (Evans, 2017). Type 2 does not always produce a normative response; it can equally well rationalize false Type 1 responses (Evans, 2017). Type 2 processes can, however, reflect on and alter Type 1 responses.

In the examples discussed, intuition was associated with poor performance, but intuitive thinking can also be powerful and accurate expert intuition. High skill and mastery get acquired by prolonged practice, and the performance of skills beomes rapid and effortless. For example, the proverbial master chess player who walks past a game and declares "White mates in three" without slowing is performing intuitively (Kahneman, 2003; Simon & Chase, 1973), as is the experienced nurse who detects subtle signs of impending heart failure (Klein, 1998; Gawande, 2002). Eye-tracking studies in chess grandmasters show that the number of moves they look ahead is only two or three (up to a maximum of five), the same as mere experts. Their visual search, however, indicates that the quality of the moves noticed is higher in grandmasters. Their perceptual abilities enable them to reproduce mid-game positions by only looking at them for a few seconds. This skill is attributed to a better intuitive understanding of the "meaning" of positions (Hastie & Dawes, 2010). Their system 1 developed a superior sense of "where to look," as Herbert Simon and William Chase (1973) pointed toward visual-perceptive processes underlying mastery. Uses of intuition are similar to rules of perception and often rely on visual analogies (Kahneman & Frederick, 2002; Kahneman, 2003).

The distinction between intuition and reasoning has been of considerable interest to psychologists (e.g., Chaiken & Trope, 1999; Gilbert, 2002; Sloman, 2002; Stanovich & West, 2002; Kahneman & Klein, 2009). As both types of processes have relative advantages, there are many other examples of tasks "fast" or "slow" thinking processes excel in, and there is substantial agreement on the characteristics that distinguish the two types of cognitive processes. Actions by type 1 processing are fast, automatic, effortless, associative, governed by habit, and difficult to control or modify. Type 2 processing is slower, serial, effortful, and deliberately controlled; they are also relatively flexible and can follow learned rules. The difference in effort is the best differentiating cue (Stanovich, 2000; Kahneman, 2003; 2011).

Dual tasks have been used in hundreds of psychological experiments to measure the attentional demands of different mental activities (Pashler, 1998; Kahneman, 2003). Studies using the dual-task method suggest that the self-monitoring function belongs to the effortful operations of System 2. People who are occupied by a demanding mental activity (e.g., attempting to hold in mind several digits) are much more likely to respond to another task by blurting out whatever comes to mind (Kahneman & Gilbert, 1989).

Stanovich, Toplak, and West (2016) also comment on the relationship between type 1 and type 2 processes and the environment. The validity of Type 1 processing depends on benign environments to provide cues that elicit adaptive behaviors. In hostile environments, reliance on heuristics can be costly (Hilton, 2003; Over, 2000; Stanovich, 2004). A benign environment contains useful cues that can be exploited by heuristics. Cues in a benign environment have high validity, and no other individuals are present who would change their behavior to exploit cues used by others. In a hostile environment, few usable cues for autonomous processes are present, and there are actively misleading cues (Kahneman & Klein, 2009). As the title "Decision making and rationality in the Modern World" (Stanovich, 2010) suggests, they categorize the information ecology of living surrounded by advertisements and other cues of the modern world as a hostile environment for type 1 processing..

3.2.3 Framing Effects

The rational agent model makes two strong assumptions about preferences and choices: extensionality (Arrow, 1982) and invariance (Tversky & Kahneman, 1986). This means preferences are independent of varying descriptions of the choice problem if the potential outcomes stay the same. These criteria are not matched by actual human choice behavior, where changing relative salience of different aspects of extensionally equivalent descriptions leads to different choices, so-called framing effects (Kahneman, 2011).

The classical thought experiment the participants got confronted with to test the rational agent model was "the Asian disease problem" (Kahneman & Tversky, 1981). A decision with the same expected value is presented in two different wordings:

"The effect of variations in framing is illustrated in problems 1 and 2.

Problem 1 [N = 1521: 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 is as follows:

- *If Program A is adopted, 200 people will be saved. [72 percent]*
- *If Program B is adopted, there is a 1/3 probability that 600 people will be saved and 2/3 probability that no people will be saved. [28 percent]*

Which of the two programs would you favor?

The majority choice in this problem is risk-averse: the prospect of certainly saving 200 lives is more attractive than a risky prospect of equal expected value, that is, a one-in-three chance of saving 600 lives.

A second group of respondents was given the cover story of problem 1 with a different formulation of the alternative programs, as follows:

Problem 2 [N = 155]:

- *If Program C is adopted 400 people will die. [22 percent]*
- *If Program D is adopted, there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die. [78 percent]*

Which of the two programs would you favor?

The majority choice in problem 2 is risk-taking: the certain death of 400 people is less acceptable than the two-in-three chance that 600 will die. The preferences in problems 1 and 2 illustrate a common pattern: choices involving gains are often risk-averse, and choices involving losses are often risk-taking. However, it is easy to see that the two problems are effectively identical. The only difference between them is that the outcomes are described in problem one by the number of lives saved and in problem two by the number of lives lost. The change is accompanied by a pronounced shift from risk aversion to risk taking."

Thomas Schelling (1984) shows in a related policy example that the intuitive preference to protect the poor can lead to drastically different decisions because of the wording. Certain outcomes are overweighted depending on how a situation is described, and the different representations of outcomes highlight and mask features.

The passive acceptance of the wording can be overcome by applying attention and mental work to actively construct a canonical representation for all extensionally equivalent descriptions. Invariance in practice needs so elaborate computations it cannot be achieved in a finite mind (Kahneman & Gilovich, 2002, Kahneman, 2003; 2011).

3.2.4 Prospect Theory

Historically formal models did not differentiate between descriptiveness and normativity, equating them with people's actual choices under risk. In contrast, prospect theory (Kahneman & Tversky, 1979) formally describes how people make decisions. It departures from predecessor choice models and is built on experimental research about human economic decision-making.

Important improvements are the conceptual movement away from the importance of "states of wealth" to "changes in wealth," which means a conceptual adaption from reference-independence to reference-dependence and different analogous evaluations of gains and losses.

In Bernoulli's theory of expected utility (1738), the utility was assigned to states of wealth, and the utility of decision outcomes was based on states of endowment. It was a reference-independent prescription to maximize states of wealth (Bernoulli, 1738) but was also understood as a description of the choices of reasonable men (Gigerenzer et al., 1989). There was no assumption of tension between normativity, prescription, and description.

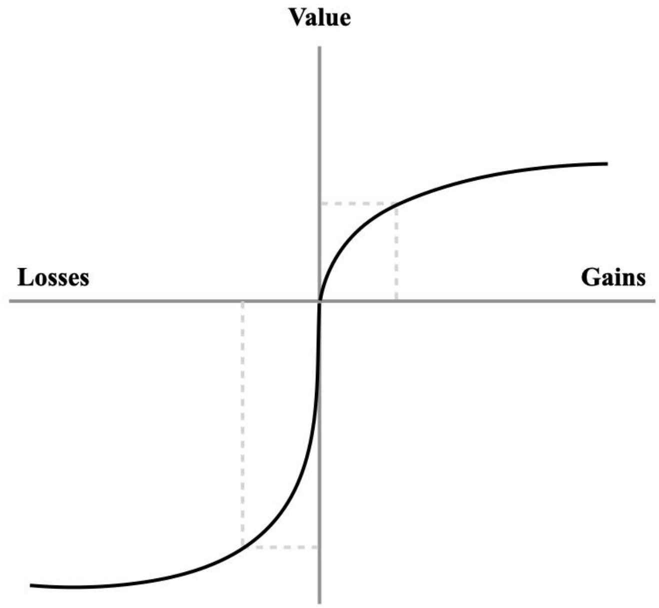

Figure 2. A schematic depiction the prospect theory, a value function for changes. It illustrates how losses from a reference point are valued by a factor of 2 - 2.5 worse than equally good gains are valued positively. Prospect theory is consistent with the marginal sensitivity of psychophysics (Tversky & Kahneman, 1984; Kahneman, 2003).

The empirical study of choice under risk lead Kahenman and Tversky (1979) to the development of prospect theory, visually represented in Figure 2. Its core idea is a value function that is kinked at the reference point and by a factor of 2 - 2.5 steeper for gains than for losses. It is concave in the gains domain and risk-aversely Conner in the domain of losses. The carrier of utility is the state change — gain or loss — from the current reference point (Kahneman et al., 1991; Kahneman & Tversky, 1992). It is consistent with the well-grounded laws of psychophysics, as the value function diminishes in its marginal sensitivity to change (Tetlock & Meller, 2002). Similarly, perceptual systems are designed to notice changes and differences and to enhance accessibility. The senses are reference-dependent, which means they notice a difference in prior context.

	GAINS	**LOSSES**
HIGH PROBABILITY Certainty Effect	95% chance to win $10,000 Fear of disappointment RISK AVERSE Accept unfavourable settlement	95% chance to lose $10,000 Hope to avoid loss RISK SEEKING Reject favourable settlement
LOW PROBABILITY Possibility Effect	5% chance to win $10,000 Hope of large gain RISK SEEKING Reject favourable settlement	5% chance to lose $10,000 Fear of large loss RISK AVERSE Accept unfavourable settlement

Figure 3. The table represents the “fourfold risk patterns”, which follows from prospect theory. It contains each a certainty effect for high probabilities and a possibility effect for low probabilities in the context of gains and losses. It makes predictions about contexts in which humans typically behave risk seeking or risk averse (Tversky & Kahneman, 1984; Kahneman, 2003).

The value function of the prospect theory is anchored at the status quo reference point. The weighting function is anchored in certainty and impossibility. On both ends, a qualitative shift occurs from impossibility to possibility or from possible to impossible. Thus the change in subjective probability from 0.0 to 0.01 or .99 to 1.0 are more consequential than any middle-range movement (Tetlock & Mellers, 2002). This leads to a four-fold pattern of risk attitudes (Kahneman & Tversky, 1986).

Effects that follow from prospect theory are the "fourfold risk patterns", depicted in Figure 3. The certainty effect typically makes humans behave risk-averse with a high probability of gains and risk-seeking in high chances of losing. The opposite effect is the possibility effect; small probabilities of winning lead to risk-seeking behavior, and in the case of a low probability of a loss, it leads to risk-averse attitudes (Kahneman & Tversky, 1979).

The discovery of the prospect theory was what earned Kahneman the Noble prize in 2003. In 2020 a group of researchers replicated the findings (Ruggeri et al., 2020) and reestablished the trust in the theory after the replication crisis in psychology and other social sciences had put many theories in doubt (Diener & Biswas-Diener, 2016; Ioannidis, 2005; Schooler, 2014; Aschwanden, 2018).

3.2.5 Axiomatic Rationality and Intelligence

Stanovich (2009; 2019) and Stanovich et al. (2016) improved the dual-process framework into a more detailed and nuanced architecture of the cognitive processes, the tripartite structure of the mind. A preliminary diagramatic representation is shown in Figure 4. The tripartite model encompasses the dual processes of cognition proposed by Kahneman (2003) while offering concrete loci for intelligence and individual differences in rationality (Stanovich & West, 2000; 2002). Based on the Tripartite Model, Stanovich et al. (2016) established a taxonomy of errors in rationality, which also allows for a more nuanced measurement of rationality, the Rationality Quotient, in multiple subtests.

Many of the examples lay people think of as intelligence, so-called permissive theories, are actually captured better by the concept of rationality in cognitive science than by what an IQ test measures in psychology. To move rationality and intelligence research closer together, Stanovich et al. (2016) transformed the dual-process framework into a tripartite structure of mind that encompasses Fluid intelligence, Crystallized Intelligence, and thinking dispositions. In their work, Stanovich et al. (2016) build on the standard grounded Cattell-Horn-Carroll (CHC) theory of intelligence (Carroll, 1993; Cattell, 1963, 1998; Horn & Cattell, 1967).

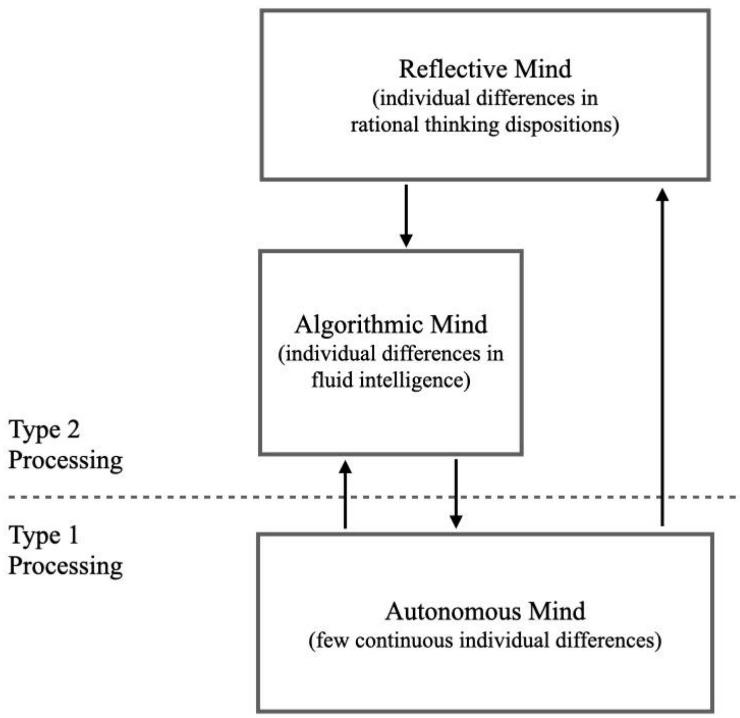

Figure 4. A preliminary depiction of the tripartite model of the mind, an advanced understanding of the dual-process theory, by Maggie E. Toplak, Richard F. West and Keith E. Stanovich (2016). It offers an understanding of loci for fluid intelligence, individual differences in rationality, and thinking dispositions and how they are all encompassed in rationality. The tripartite model of the mind allows for a detailed analysis of error causes in tasks that test for axiomatic rationality. An additional, more fleshed out, and nuanced version can be found in Stanovich et al. (2016.

The main difference between rationality and intelligence is seen in typical or optimal performance situations (Ackerman & Kanfer, 2004; Cronbach, 1949; Sternberg et al., 2008). Optimal, or maximal, performance situations define a task at hand externally to require maximum cognitive performance, as is usual in tests of intelligence or cognitive aptitude. They test the capabilities and individual differences of the algorithmic mind. Typical performance situations, however, are closer to everyday situations. Here the decision of how effortfully and with which algorithm to perform the task is internally decided by the reflective mind of the solver. While individual differences in thinking dispositions or "cognitive styles" are located in the reflected mind, testing for how effortful the person thinks on average. For example, the average time of thinking before responding, the tendency to collect information before making up one's mind, and the average depth of computation.

Actual measures of intelligence in use assess only algorithmic mind cognitive capacity, while the whole tripartite structure of the mind, including thinking dispositions, are assessed with cognitive reflection tasks (Toplak et al., 2011). To successfully answer a cognitive reflection task multiple steps have to be completed successfully: preattentive processes of Type 2 have to detect a conflict (Evans, 2008), the Type 1 response has to be successfully interrupted and inhibited (a hallmark of executive functioning; Best et al., 2009; Hasher et al., 2007; Miyake & Friedman, 2012), a better response as a substitute for the automatic reaction has to be generated and the wrong type 1 response has to be overwritten (Evans, 2007, 2010; Evans & Stanovich, 2013). These steps have to be done while maintaining cognitive decoupling from the environment to not confuse simulations and reality (Stanovich, 2011), a major contributor to the distinct property of seriality in Type 2 (Duncan et al., 2008; Gray et al., 2003; Hicks et al., 2015; Kane, 2003; Kane & Engle, 2002; Kane et al., 2005; Salthouse, Atkinson, & Berish, 2003). For the individual differences in rational thinking, therefore both fluid intelligence and the reflective mind, thinking dispositions, are significant (Stanovich et al., 2016).

The Taxonomy of Cognitive Errors.

By analyzing successful and unsuccessful solution processes in cognitive-reflection tasks with the tripartite model, Stanovich et al. (2016) could identify different error sources. This results in a taxonomy of mistakes and their causes. The two main dimension causes are either, first, miserly processing, which means failing to adequately activate Type 2 processing, or, second, mindware problems.

Humans are cognitive misers because their basic tendency is to default to processing mechanisms of low computational expense (Dawes, 1976; Kahneman, 2011; Simon, 1955, 1956; Taylor, 1981; Tversky & Kahneman, 1974; Stanovich, 2004, 2009). Cognitive effort is a constant trade-off between power and expense (Tomlin et al., 2015), as great concentration and attention are often experienced as aversive (Kahneman, 1973; Kurzban, Duckworth, et al., 2013; Navon, 1989; Westbrook & Braver, 2015). The miserly choice does not necessarily imply only a tendency to Type 1 processing; it can also be to choose the less demanding of multiple type 2 processes. In the case of defaulting to the autonomous mind in cognitive reflection tasks, no struggle or conflict is either happening or even noticed; an override failure occurs when a conflict is detected but not won by a Type 2 override. In economics or psychology considered to reflect problems of self-control (Baumeister & Vohs, 2003, 2007; Hofmann, Friese, & Strack, 2009; Loewenstein, Read, & Baumeister, 2003).

Sloman (1996) talks here about the "quintessential dual-process conflict" by quoting Stephen Gould's introspection about the Linda problem, which tests for the conjunction fallacy: "I know the [conjunction] is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me - 'but she can't be a bank teller; read the description'" (Gould, 1991, p. 469)

In defaulting to serial associative cognition, type 2 processing is used, but to such a limited extent that it has an inappropriate focal bias, for example, making expected value calculations but passively accepting the given formulation in framing problems (Kahneman, 2003). Kahneman (2011) calls this the "WYSIATI tendency" ("What you see is all there is"). Information that is systematically unavailable is not actively made available by considering counterfactuals. The decision is influenced by the environment of the moment.

The fourth error category is about mindware gaps. In this error category, the subject responds with a wrong and intuitive System 1 response. No conflict or override is necessary as the subject has not acquired the mindware to have a conflict between an intuitive and normative response. Typical areas of lacking knowledge are probabilistic reasoning, causal reasoning, logic, and scientific thinking.

Mindware can also cause errors, where an override happens, but the acquired mindware is contaminated and the direct cause of a wrong choice. In their book, Stanovich et al. (2016) refer to the very extreme example where Nazi officers admitted any civilized human would experience a moral disgust reaction to committing genocide, that, however, what they are about to do, is "the right thing" defined by "a higher goal." Their ideology is the mindware used to override their type 1 response of moral disgust.

Through the tripartite model of the mind, as a more detailed understanding of the dual-process theory, it was possible for Stanovich et al. (2016) to understand the interplay between intelligence and rationality and to derive a taxonomy of error sources in cognitive-reflection tasks as an analogy for real-life situations. These error sources are based on miserly processing and mindware problems (Stanovich et al., 2016; Stanovich, 2018).

Testing for the Rationality Quotient.

With the development of a detailed error taxonomy, Stanovich made it possible to develop the "Comprehensive Assessment of Rational Thinking" (CART) test, later called the "Rationality Quotient" (Stanovich, 2016; Stanovich et al., 2016). In direct relation to the dimensions of the established error taxonomy, the subtests of the CART can be categorized into tasks with high processing requirements, tasks high in knowledge requirements, tasks testing for the avoidance of contaminated mindware, and tasks testing for different thinking dispositions.

The CART uses the prodigious number of tasks that measure components of rational thinking assembled over the last decades (Baron, 2008; Kahneman, 2011; Stanovich, 2011). Stanovich et al. (2016) refuted training making people better at solving the CART as a cause for concern about its validity. They write that the conception of a rationality quotient is grounded both in thinking dispositions and the acquisition of certain mindware. It can straightforwardly follow that teaching or understanding the mindware by taking the test questions can be a means to increase the knowledge and use of the mindware.

This conceptualization of the CART and its subtests is, as explained before, built on the axiomatic framework. It says that if people's preferences follow transitivity, independence, and other axioms of choice, and show descriptive invariance (Tversky & Kahneman, 1986), they behave as if they are maximizing utility. This is what makes people's degree of rationality measurable by using the count of failed tasks as the inverse measure of rationality. As a bottom line, the CART assesses how often the axioms of choice are violated.

Both Chapters Three and Four aim to transport foundational understanding of the axiomatic and ecological approach to rationality and are written to avoid typical past misunderstandings of debates around the topic. In case of robust foundational knowledge of the literature, the reader may continue directly to Chapter Five.

This concludes Chapter Three, giving an overview of the axiomatic approach and the empirical work done on its ground. The following Chapter – Chapter Four – will continue to overview the ecological approach to rationality.

4. Ecological Rationality

Since the 1990s, an alternative interpretation of the findings from the heuristics and biases research program has been brought forward (Gigerenzer, 1991; Gigerenzer et al., 1999; Smith, 2003, 2008; Dekker & Remic, 2019). Contributing to this alternative interpretation have been evolutionary psychologists, adaptationist modelers, and ecological theorists (Anderson, 1990; Cosmides & Tooby, 1996; Gigerenzer, 2007; Oaksford & Chater, 2007; Todd & Gigerenzer, 2000).

They took the modal response in most of the classic heuristics and biases experiments as indicating an optimal information processing adaptation on the part of the subjects. Instead of questioning human rationality, they proceeded to question and reinterpret the normative theories used as benchmarks. This group of theorists—who argue that an assumption of maximal human rationality is the proper default position to take—have been termed the Panglossians (Stanovich, 2010).

In contrast to axiomatic rationality, where consistency with the axioms of rationality is the measure of rationality, this stream of research is termed ecological rationality, where the adaptionist value and validity of action to the agent's environment in a specific context became the focus. Ecological rationality asks how decision mechanisms can produce useful inferences by exploiting the structure of information in their environment.

Accordingly, there is a difference in what the different schools deem rational. While the violations of the rational choice theory are understood as an indication of irrationality from an axiomatic perspective, the ecological rationality notion questions the normative validity of the axioms of rationality. Moreover, it interprets the empirical findings in an alternative fashion. Instead of questioning human rationality, it questioned the validity of the normative notions.

From a distance, the divide between the different forms of rationality is sometimes portrayed as the axiomatic or Meliorist side focusing on the errors in human cognition (biases). At the same time, the ecological approach points out the adaption value of these simple rules of thumb (heuristics). While this is a false dichotomy — the heuristics and biases program explicitly states that heuristics are the optimal choice in most cases — it can be a valuable shortcut to knowing which positions are associated with the different approaches.

This Chapter on the ecological approach to rationality contains multiple parts, roughly following the advancements made by the tradition in chronological order.

First, it will lay out the initial fast-and-frugal heuristics approach to ecological rationality and introduce the adaptive toolbox (Gigerenzer & Goldstein, 1996; Gigerenzer & Todd, 1999; Martignon & Schmitt, 1999; Gigerenzer et al., 2011; Raab & Gigerenzer, 2015; Gigerenzer, 2015; Hafenbrädl, 2016). This section follows the argument of why biased minds make better inferences (Gigerenzer & Brighton, 2009) by describing the less-is-more effect. It rests on the argumentation of a bias bias in behavioral economics and understanding errors through a bias-variance decomposition (Brighton & Gigerenzer, 2015; Gigerenzer, 2018).

The next part will give an overview of theoretical developments regarding the statistical foundations of ecological rationality in comparison to the axiomatic approach (Gigerenzer, 2019; Brighton, 2018, 2020) as well as the underlying interpretation of uncertainty (Kozyreva &

Hertwig, 2019; Kozyreva et al., 2019; Hertwig et al., 2019, Kay & King, 2020). Should the treatment of some concepts and theories from ecologial rationality appear superficial, then this is to be attributed to them being prominently featured in the central argumentation of this thesis in the upcoming Chapter Five.

4.1 Simple Heuristics

Historically, the ecological approach to rationality took a big step back to ask with a beginner's mind how humans and other animals cope effectively with the uncertainty of the natural world and what, after reformulating the rationality question, constitutes rational decisions in the first place.

Ecological rationality answers these questions by determining how cognitive mechanisms can generate instrumental inferences, which take advantage of the structure of how information is represented in the environment (Todd et al., 2000). In this context, the focus of ecological rationality on simple heuristics from its foundation onwards quickly becomes apparent. While the axiomatic approach acknowledges the functionality of heuristics in many cases, the centrality of simple heuristics in ecological rationality rests on different conceptual foundations and assumptions. Moreover, heuristics do not work despite being limited; their functional value stems precisely from their simplicity and limitations (Gigerenzer & Goldstein, 1996; Gigerenzer & Todd, 1999).

4.1.1 The Fast and Frugal Approach to Ecological Rationality

The so-called "fast and frugal" approach proposes ecologically rational simple heuristics in three steps; simple search, stopping, and decision rules. The heuristics include simple rules to control the search for information in the environment or memory, stopping rules regarding the search processes, and decision rules regarding how to use the information found to make judgments and decisions (Gigerenzer & Todd, 1999; Todd et al., 2000; Raab & Gigerenzer, 2015; Hertwig et al., 2019).This class of heuristic includes take-the-best (Gigerenzer & Goldstein, 1996), fast and frugal trees (Martignon et al., 2008), and the priority heuristic (Brandstätter et al., 2006; Drechsler et al., 2014).

Ecological rationality asks whether heuristics are valid for consequentialist or means-end rationality in a specific environmental structure. It can be accurately defined by describing the information structures in environments and the heuristics adapted to make good decisions (Todd et al., 2000; Schurz & Hertwig, 2019).

For example, in environments where cues are typically non-compensatory and declining rapidly in their relative validity, the take-the-best heuristic is ecologically rational (Martignon & Hoffrage, 1999; Todd et al., 2000).

The take-the-best heuristic makes choices between available options by searching through available cues in order of their validity in the environment, then stopping once the first cue is found that distinguishes the options, and then choosing the first option indicated through a higher cue value (Gigerenzer & Goldstein, 1996; Todd et al., 2000). The logic of the rule is that it bases its decision on only the one most valid and therefore best cue and ignores any further information. Significant inspiration for this ecological performance classification stems from the ideal observer analysis and signal detection theory, as the attention toward cues indicates.

4.1.2 The Adaptive Toolbox

Challenging the axiomatic or "Olympian" ideal of rationality is a central pillar of the ecological approach. Humans do not master the uncertainty surrounding them as if they were reducing it to quantifiable risk (Hertwig et al., 2019). Instead, humans deploy a range of cognitive tools, referred to as the mind's adaptive toolbox (Gigerenzer, 2002; Hertwig et al., 2019; Mohnert et al., 2019). Cognitive tools in the adaptive toolbox are simple heuristics (Todd et al., 2012; Arkes et al., 2006), search strategies, and crowd aggregation rules.

Martignon et al. (2003, 2008) introduced the concept of fast-and-frugal (decision) trees based on fast and frugal decision rules. They are simple graphical structures that categorize objects by asking multiple binary categorization questions with one question, along the dimension of a high-value cue, at a time. Fast-and-frugal decision trees constitute lexicographic heuristics or linear classification models with non-compensatory weights and thresholds (Gigerenzer et al., 2011). These decision trees' performance, operationalized as accuracy and robustness, approximates Bayesian benchmarks (Laskey & Martignon, 2014).

While there is no perfect axiomatic normative ideal of formal requirements to determine ecological rationality, one can determine the relative values of decision rules along multiple dimensions. Dimensions along which heuristics are valued are extended beyond purely formal and epistemic goals to, among others, the accuracy of judgment, speed, and frugality. Frugality here means how many cues need to be searched before coming to a reasonable decision. The normativity of ecological rationality stems from using heuristics that have a higher value in a specific environment. Accordingly the normativity is relative and context-dependent, instead of formal and absolute (Gigerenzer & Goldstein, 1996; Todd et al., 2000; Gigerenzer & Gaissmaier, 2011; Todd & Gigerenzer, 2012; Gigerenzer, 2019).

4.1.3 Homo heuristics, the Less-is-More Effect, and the Bias Bias

The ecological rationality research tradition argues that simple heuristics more accurately approximate and embody Herbert Simon's bounded rationality. They enable organisms and other systems to make better choices while being internally and externally constrained. Behavioral economics and the heuristics and biases tradition emphasize internal limitations, such as limited memory and computational power. By the end of the 20th century, this led to the wrong idea that more computational resources would free a cognitive system from relying on heuristics. Hertwig and Gigerenzer (2009) summarize the misconceptions as "1. Heuristics are always second-best. 2. We use heuristics only because of our cognitive limitations. 3. More information, more computation, and more time would always be better."

This notion leaves the second part of the constraints of cognitive systems out of the picture, the uncertainty of the external world and the constraints it imposes on decision-makers (Todd et al., 2000; Gigerenzer, 2000). Ecological rationality argues that heuristics make better inferences and are superior to more extensive computation, not despite but precisely because they ignore information (Dawes, 1979; Martignon & Schmitt, 1999; Gigerenzer & Todd, 1999; Brighton &

Gigerenzer, 2009; Gigerenzer & Gaissmaier, 2013). Systematic investigations even challenged long-standing arguments of simple rules failing in competitive interactions, instead heuristics turned out to be adapted well to game theoretic standoffs (Spiliopoulos & Hertwig, 2020).

Part of the judgment error regarding the functionality of heuristics is the wrong assumption of a linear relationship between perceived problem complexity and the required complexity of the mechanism to confront it. No linear relationship exists (Todd, 2000; Stanovich, 2013; Brighton & Gigerenzer, 2015). The rationale for biased minds making better decisions is developed from the conception of the Less-is-More effect and a bias bias in behavioral economics (Gigerenzer & Goldstein, 2002; Martignon & Hoffrage, 2002; Hastie et al., 2009; Brighton & Gigerenzer, 2009).

One example that helps move from the axiomatic view of rationality to that of ecological rationality is to consider how a baseball outfielder successfully catches a ball (Brighton & Gigerenzer, 2009):

Analyzing the scene from an omniscient and omnipotent "view from nowhere," the complex problem requires equally complex mental algorithms. The outfielder "behaves as if he had solved a set of differential equations in predicting the trajectory of the ball... At some subconscious level, something functionally equivalent to the mathematical calculations is going on" (Dawkins, 1989, p. 96).

The inserted "as if" indicates Dawkins' doubts regarding whether brains perform these computations. Brighton and Gigerenzer (2007, 2009) argue that experiments have shown that it does not do so; instead, players rely on several heuristics, most and foremost the "gaze heuristic" as the simplest.

The simple rule of the gaze heuristic is that once the ball is up in the air, fix gaze on the ball, run in the direction the ball is flying, and adjust the running speed so that the gaze angle remains constant.

With the focus on one variable — keeping the gaze angle constant while running — the player ends up in the location as the ball comes down. There is no need to compute differential equations to determine the exact spot, and one can safely ignore variables like initial distance, velocity, air resistance, speed and direction of wind, and spin, among others (Gigerenzer, 2007; Brighton & Gigerenzer, 2009).

The axioms of rationality judge — along dimensions of logical consistency, coherence, and an effort to incorporate all available pieces of information — the construction of accurate and general representations of the world (Todd et al., 2000). Kay and King (2020) speak of the beauty of the Savage axioms and probability formalisms as seductively elegant and intrinsically rewarding to a potentially dangerous degree. In this manner, it was extensively and formally argued how more information is inherently superior both in theory (Carnap, 1947; Good, 1967) and in optimization-under-constraints theories for search in the world or memory (e.g., Stigler, 1961; Anderson, 1991).

Brighton & Gigerenzer (2009) point out that this is incorrect; more information is not inherently better. A trade-off between accuracy and effort is the fundamental assumption from which the value of more information follows (Heitz, 2014; Fudenberg et al., 2018), but more information or computation can decrease accuracy in increased uncertainty conditions. Therefore, simple heuristics can be more accurate than strategies that use more information and time (Gigerenzer, 2007; Brighton & Gigerenzer, 2009; Spiliopoulos & Hertwig, 2019). The ability of less information to be more valuable has been coined the less-is-more effect (Brighton & Gigerenzer, 2009).

The second argument for the robust value of heuristics is the bias-variance decomposition of errors (Geman et al. 1992). Including variance in evaluating the error of cognitive strategies can overcome the bias bias of behavioral economics (Gigerenzer, 1991; Brighton & Gigerenzer, 2009; Brighton & Gigerenzer, 2015; Gigerenzer, 2018). In conditions of radical uncertainty, cognitive systems must successfully cope not (only) with a speed-accuracy trade-off but also with the efficiency-robustness trade-off. The main argumentation of Chapter 5 will treat this line of thought in adequate detail.

4.2 Statistical Foundations, Optimality and Uncertainty in Ecological Rationality

Ecological rationality contributed to multiple significant advancements. One contribution was replacing uncertainty as a quantifiable or numerically expressable property of the environment. The positions regarding quantifying uncertainty denote a crucial difference between axiomatic and ecological rationality. The new statistical foundations of ecological rationality limit quantified risk to use in well-posed problems and statistical "small worlds." It started to conceive of uncertainty as a property of the organism-environment system (Brighton & Gigerenzer, 2012; Brighton, 2018, 2020; Felin & Koendrick, 2021; Gigerenzer, 2019; Hertwig et al., 2019; Kay &

King, 2020; Kozyreva & Hertwig, 2019; Taleb, 2017; Vale et al., 2021; Okasha & Binmore, 2012; Okasha, 2016).

4.2.1 Rationality without Optimality

Humans, however, are not confronted with quantifiable risk in their natural world. Accordingly, consistency with the axioms of rationality is only valid in statistical "small worlds," where risks are quantifiable and optimal solutions computable. Instead, humans live in a non-stationary world. As a result, they must cope with radical and unmeasurable uncertainty (Kozyreva & Hertwig, 2019; Kay & King, 2020). This ill-posed reality of statistical "large worlds" in which humans reside is known as Knightian uncertainty (Okasha & Binmore, 2012; Okasha, 2016; Gigerenzer, 2019; Kay & King, 2020; Brighton, 2020).

Historically, the different statistical requirements needed in well-posed problems or "small worlds" relative to ill-posed problems or "large worlds" has been repeatedly acknowledged (O'Sullivan, 1986; Brighton & Gigerenzer, 2012; Brighon, 2018, 2020; Gigerenzer, 2019). Remarks of that kind were made by Simon (1973), Keynes (1937), Luce and Raiffa (1957), and Marr (1973), while Savage, the founding father of Bayesian decision theory himself, considered the application of his methodology outside of small worlds "utterly ridiculous" (Savage, 1954). None of this functions as an argument against optimality modeling; instead, it is to be understood as an argument for understanding the limits of the methodology (Brighon, 2020; Kay & King, 2020).

Brighton (2018, 2022) notes the different faces of the same interdisciplinary orthodoxy of optimality. Examples are optimal Bayesian decision-makers in the cognitive sciences, optimal foragers in biology, and Bayesian maximizers of expected utility in economics

As an alternative approach — rationality without optimality — Brighton (2018, 2020) points to the two cultures of statistical modeling (Breiman, 2001). In statistical modeling, the distinction between data modeling and algorithmic modeling directly reflects the large approximations made by ecological and orthodox rationality.

Analogous to Bayesian optimality modeling, data modeling views the conjecture of a stochastic model of nature's black box as an essential step in statistical inference. Alternatively, algorithmic modeling, analogous to ecological rationality, is an incremental search for learning algorithms that can, to varying degrees, accurately predict the input-output relationship. This clash reflects long-standing disagreements in statistics (Breiman, 2001; Brighton, 2018, 2020).

4.2.2 Uncertainty as a Property of the Organism-Environment System

In the ongoing ecological rationality endeavor, an important question is identifying a good fit between tools and environments. This focus rests on a systemic view of ecological uncertainty (Kozyreva & Hertwig, 2019). No longer is uncertainty attributed to the state of mind (epistemic uncertainty as the subjective degree of belief) or the environment (aleatory uncertainty as stable frequencies displayed by chance).

Instead, uncertainty is seen as an interactive and relational property of the organism-environment system (Kozyreva & Hertwig, 2019; Hertwig et al., 2019). This system of organism and environment together is sometimes approximated by the analogy of Herbert Simon's scissors metaphor of bounded rationality. In the metaphor, one blade represents the internal constraints of the organism and the other that of the external system (Brunswik, 1955; Viale, 2021).

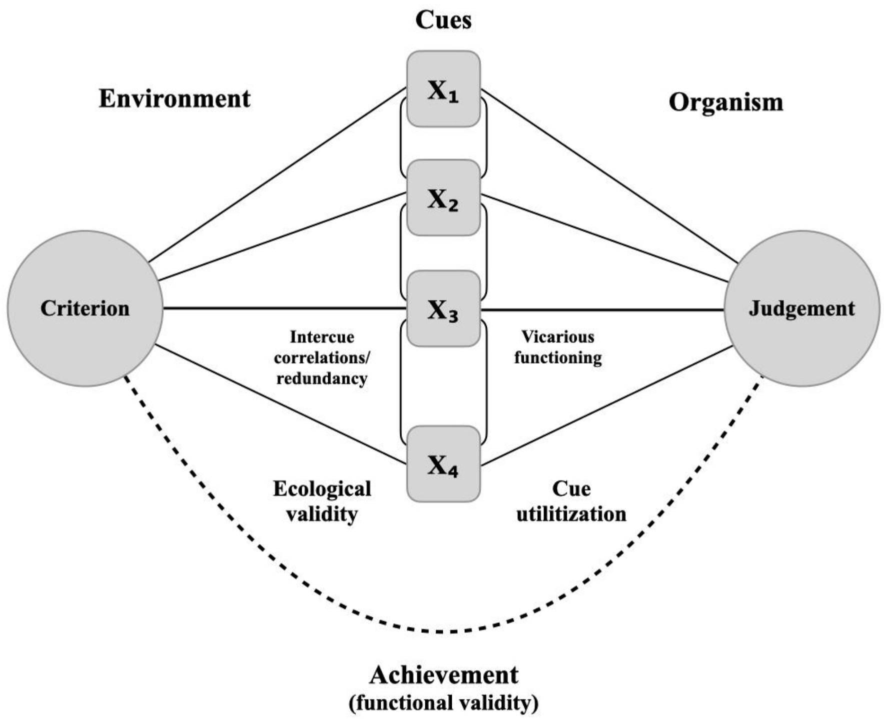

Figure 5. Adapted visual depiction of Brunswik's lens model framework for judgment. It illustrates the notion of uncertainty as a property of the organism-environment system in its semiotic discourse. The representation is based on Brunswik (1952), Dhami et al. (2004), Hastie and Dawes (2010), and Kozyreva and Hertwig (2019).

The notion of the environment-organism system is often explored through Brunswik's lens model, which visualizes two critical properties of this systemic notion of uncertainty. It is depicted in Figure 5.

Firstly, the distance of uncertainty in the relationship between causes and effects and, secondly, the underpinning inferential nature of human judgment and reasoning. In it, the process of cognition is ordered around two focal variables, and its temporal inference process is mediated by cues (Kozyreva & Hertwig, 2019, Brunswik, 1955). As the lens model's first part, the environment explores relations between available proximal cues and distal objects and criteria. The organism as the second part focuses on how the organism utilizes the cues to achieve its goals. No one-to-one relation between causes and effects or distal criterion and proximal cues exists. Instead, the organism-environment system follows the principle of mutual substitutability, or "vicarious functioning," of means for each other (Brunswik, 1955).

Therefore, understanding uncertainty in judgment and decision-making implies accounting for how people make inferences and the characteristics of their environments. Selecting and using cues in an ongoing process is essentially the achievement of the organism necessary to forward its goals in the world (Kozyreva & Hertwig, 2019).

This concludes Chapter Four, giving an overview of ecological rationality, including the robust beauty of simple heuristics, the adaptive toolbox, and the statistical foundations of ecological rationality in radical uncertainty. Both Chapters Three and Four aimed to transport foundational understanding of the axiomatic and ecological approach to rationality. They are written to avoid typical past misunderstandings of debates around the topic. The following Chapter – Chapter Five – will present this thesis's central strain of argumentation.

5. Triangulating Rationality Between Optimality and Uncertainty

In the literature, rationality is often used synonymously with optimal judgment and decision-making under uncertainty, or simply put, "doing the right thing" (Kahneman & Tversky, 1974; Hastie & Dawes, 2010, Russell & Wefald, 2003). The underlying uncertainty is conceived here as risk and accordingly is probabilistically quantifiable, as is also typical of economic analysis. A correct assessment of risks and weighing of alternatives of action is what is considered to be rational behavior, in line with the normativity of logic, probability calculus, and Bayesian decision theory (Hertwig et al., 2019). In this conception, the deviation from rationality – the extent of given irrationality – can be measured as the distance of deviation from theoretical normative ideals. In recent years, the notion of uncertainty as something that can be measured has come under fire in multiple disciplines (Brighton, 2020; Felin & Koendrick, 2021; Gigerenzer, 2019; Hertwig, Pleskac, & Pachur, 2019; Kay & King, 2020; Kozyreva & Hertwig, 2019; Taleb, 2017; Vale et al., 2020). The economists Kay and King (2020) argue that humans are actually confronted with radical and unmeasurable uncertainty in the non-stationary world they inhabit, also known as *Knightian uncertainty*. They point out that the appliance of mathematical models and probability theory outside of small, abstracted, or stationary worlds in the real world exceeds the validity of the methods; an aspect already criticized by Keynes and the founding fathers of said methods.

Similar notions of immeasurability of uncertainty were recently developed in the ecological critique of "orthodox" axiomatic rationality. A major conceptual move went beyond uncertainty as a measurable property of the environment to a systemic view of the uncertainty as a property of the organism-environment system (Kozyreva & Hertwig, 2019). One analogy used to describe this system was that of a married couple between organism and environment by Brunswik, and another one the scissors metaphor by Simon, where one blade represents the internal constraints of the organism and the other that of the external system (Brunswik, 1955; Viale, 2021). The response by the ecological approach to this conception of uncertainty is to offer a taxonomy of tools specific to the given environment. However, organisms will never be omniscient and have a priori insight into the properties of their environment (Felin et al., 2017; Wimsatt, 2007). This thesis argues that while having tools for different environmental properties is valuable, being able to assess and determine what is relevant about a given situation comes first and is the most crucial cognitive skill in rationality. This process of problem reframing and sense-making is also referred to as abductive reasoning.

Precisely the ongoing, procedural, and dynamical cognitive transformation of an uncertain environment – or "large world" (Savage, 1954) – into an abstracted, tractable "small world" is at the center of the phenomenon of rationality. Consequently, rationality is about continuously solving the frame problem, in line with Dennett (1984), and refer to this emerging bottom-up transformation of the environmental representation by negotiating different cognitive conflicts as 'relevance realization' (Vervaeke et al., 2012). Relevance realization brings together empirical prescriptive research from the axiomatic approach to rationality, like active open-mindedness and mental decoupling (Stanovich, West, & Toplak, 2016), with the ecological conception of uncertainty and the selection of adaptive tools (Hertwig, Pleskac, & Pachur, 2019), and also accounting for the insight problems of the naturalistic decision-making approach (Klein, 2013).

The upcoming line of argumentation is related to general developments in cognitive science, away from logicist computationalism and propositional cognition, and towards the 4E-approach (Rolla, 2019) of viewing cognition through embodied dynamism. This line of argumentation has implications for categorization and perception (Hofstadter, 2011; Felin, 2021, Hoffmann, 2015), philosophy of science (Thagard, 2012; Hartmann & Sprenger, 2019), economic theory (Kay & King, 2020), as well as machine learning and applied mathematics (Weinan, 2021).

This thesis elaborates in-depth how the debate between axiomatic rationality or optimality modeling on one side and ecological rationality on the other can be reexamined from the angle of understanding the rationality problem as solving the frame problem. The two approaches are not genuine opposites, but rather complementary to one another, studying the same phenomenon from different perspectives. This work aims to differentiate the knowledge about the rationality phenomenon generated from different perspectives from the artifactual disagreements that stem as by-products from the different methods, perspectives, and assumptions used (Giere, 2006; Wimsatt, 2007; Callebaut, 2012; Massimi & McCoy, 2019). As a process such triangulation can enrich each approach by discriminating between the perspectival disagreement and the fundamental gap in knowledge integration.

5.1 Perspectives on Rationality

This thesis takes the stance that not all relevant advancements each tradition highlights have been integrated into each other. As a result, there is further room to move them closer together to a shared understanding of rationality. This work addresses exactly this gap and brings forth insights that are not reducible to the models of one or the other party, instead it asks which parts of the debate are rooted in insights of the respective other side that have not yet been integrated. The argumentation differentiates between the different perspectives on the phenomenon of rationality and how to distinguish them from parts that are purely perspectival artifacts.[2]

The descriptive psychological research of the heuristics and biases tradition was built on the axioms of rationality from theoretical economics and decision science. This line of research attempts to identify heuristics and the biases – systematic deviations from rationality – they cause. Prescriptive attempts were made on this foundation; recommendations on how to decrease deviations from the normative ideal of axiomatic rationality (assuming this gap is essential to the Meliorist view). However, these prescriptions (e.g. Stanovich et al., 2016) ignore multiple problems that ultimately lead back to questioning the theoretical foundations of optimality modeling. These prescriptions usually presuppose a cost-free world: when taking both thinking effort and opportunity costs of delayed action into account, the biases are often Bayes optimal and accordingly cannot be naively improved upon. Whether or not this is the case can be approximated by resource-optimal analysis and modeling (Lieder and Griffith, 2019). The idea of computational rationality (Gershman et al., 2015) needs to be introduced as an underpinning. This notion integrates internal and external computational constraints on an

[2] In Wimsattian scientific perspectivism and robustness analysis, the crucial decision is to distinguish between what is real and what is illusory, ergo the object of focus from artifacts of perspective. Even the best scientific models have artificial assumptions and when observing results, one has to ask whether they depend on the essentials of a model or on the details of the simplifying assumptions. Wimsatt himself tells the story of how the idea of group selection was wrongly refuted, because the absence of group selection effects resulted from the assumptions in the modeling approaches (e.g. random mating within populations) (Wimsatt, 2007).

information-processing system embedded in an online (having a temporal dimension) environment.

On an even more fundamental level, according to the no-free-lunch theorems by Wolpert (1995; 1997), this means there cannot be one single search algorithm optimal for all cases independently of the search space it is applied to. From this, it suggests that there cannot straightforwardly be one prescription to think more rationally independent of the environment. Instead of assuming a mere speed-accuracy trade-off that implies a more effortful process to always be better while implying more costs, this work argues for conceptualizing the trade-off additionally along another dimension, namely between efficiency and robustness. Underlying is a shift of paradigm from cognitivism to embodied dynamicism, taking a more adaptionist view of ecological rationality. Both conflicts have to be continuously solved by systematically overcoming the frame problem of what is relevant in a given embedded and situated system of organism and environment. To understand what this means, the concept of relevance realization and its emerging framework by Vervaeke, Lillicrap, and Richards (2012) will be introduced. This is necessary to understand how introducing the concept of computational rationality and bounded optimality constrain the bounds of the rationality debate on all sides.

5.2 Conceptual Robustness Analysis

Scientific perspectivism, a stance on the philosophy of science taken prominently by Giere, Van Fraassen, and Wimsatt is one of the most potent paths of the current cultural repertoire to approach philosophical issues applied in the understanding of complex, multilevel and multiscale phenomena (Callebaut, 2012; Massimi & McCoy, 2019). According to this position, science cannot as a matter of principle transcend our human perspective. One of the tools of scientific perspectivism is robustness analysis, as well as bootstrapping and triangulating conceptual models (Wimsatt, 2007). The central goal of a robustness analysis according to Wimsatt (2007), is the distinguishing of the real from the illusory; the phenomenon under study from any byproducts of modeling assumptions or perspective. This rests on the assumption that adding multiple perspectives or redundancy in observation increases reliability. Parallel organization is a fundamental principle of organic design and robust engineering. The classic paper by von Neumann (1956) made the same point on building reliable automata with unreliable components. The use of a multitude of methods to determinate and "triangulate" the existence and properties of an object of research has a long history in science. Aristotle, for example, valued having multiple explanations of world phenomena, while in Galileo's work – and those following him – the notion of robustness is implicit in the distinction between primary and secondary qualities. It is one of several conceptions involved in Whewell's method of the "consilience of inductions" (Laudan, 1971) and is to be found in several pieces of Peirce (Wimsatt, 2007).

As an illustrating example, Wimsatt (2007) chose the methods of analyzing satellite pictures of Mars, as described by Campbell (1977) in the unpublished William James Lectures. Part of the recommended procedure to understand satellite pictures involves combining multiple images. As a result, the noise, being random, averages out; but the signal, no matter how weak, adds up in intensity until it stands out enough to be recognized as such. Von Neumann (1956) used the same concept in his notion of "majority organs"; multiple parallel signals are brought together with a threshold, the signal is multiplied, while the noise clears out. Wimsatt (2007) describes common features of robustness and of robustness analysis built on these conceptions and continues to list different processes that apply these. The one most closely resembling the attempt undertaken in this work is the examination of different theoretical descriptions of the same phenomenon for matches and mismatches. This generates new hypotheses, while also modifying and refining the theories examined (Wimsatt, 1976, 1979).

In line with Wimmsattian robustness analysis, this work attempts to unify the axiomatic and ecological approaches to rationality. And while the attempt ultimately fails – no unification is possible – it does so in a productive manner. I argue that the existing gap, which can ultimately not be closed, is, following the argumentation, not based on disagreements about phenomena in the world, but rather a result of the perspectives taken to study them. Undertaking a unification, however, can enrich each perspective and distinguish the perspectival disagreement from the fundamental gap.

The most central angle to start in this argumentation is the assumption of a cost-free world in axiomatic rationality. However, when taking both thinking effort and opportunity costs of delayed action into account, the supposed biases often turn out to be Bayes-optimal and can thus not be naively improved upon (Gershman et al., 2015). Whether or not this is the case can be approximated by resource-optimal analysis and modeling (Lieder and Griffith, 2019). This work will explore this lowered normative ceiling of resource rational optimality in the next section and develop the argument from there.

5.3 Lowering the Normative Ceiling with Computational Rationality

"Imagine driving down the highway on your way to give an important presentation, when suddenly you see a traffic jam looming ahead. In the next few seconds, you have to decide whether to stay on your current route or take the upcoming exit—the last one for several miles—all while your head is swimming with thoughts about your forthcoming event. In one sense, this problem is simple: Choose the path with the highest probability of getting you to your event on time. However, at best you can implement this solution only approximately: Evaluating the full branching tree of possible futures with high uncertainty about what lies ahead is likely to be infeasible, and you may consider only a few of the vast space of possibilities, given the urgency of the decision and your divided attention. How best to make this calculation? Should you make a snap decision on the basis of what you see right now, or explicitly try to imagine the next several miles of each route? Perhaps you should stop thinking about your presentation to focus more on this choice, or maybe even pull over so you can think without having to worry about your driving?" (Gershman et al. 2015)

This example demonstrates several themes in the study of (artificial) intelligence and (computational) rationality that are usually ignored when assuming perfect (or non-bounded) rationality. These problems are also neglected in prescriptions built on studies of human judgment and decision-making based on the axiomatic approach to rationality. First of all, real-life decisions have to be made under high uncertainty and without having the full branching tree of possible futures. Even the choice of things in the environment or the own mind to pay attention to happens under radical uncertainty. Additionally, this happens often under high urgency and time pressure as well as divided attention. These factors lead to a full set of internal decision problems, like how much to think, what further information to gather to decrease the uncertainty, how far to plan ahead, and even what to think about. In a more formalized manner, these themes include the general-purpose ideal for decision-making under uncertainty, which is maximizing some measure of expected utility, under internal and external constraints. Since this is nontrivial for most real-world problems, approximations become necessary. The choice of how best to approximate becomes a decision that is subject to the expected utility calculus itself due to the introduction of bounded rationality, which highlights the limitations and high costs of the computational power of thinking.

The concepts and ideas underlying computational rationality in their historical order arose from the theory of computability, in which AI has its roots, which was developed in the 1930s. The idea of using computation to rebuild human intelligence arose. Then researchers considered the application of probability and Bayesian updating in learning, reasoning, and action with foundational arguments for probability as a sufficient measure of judging and reevaluating the plausibility of events based on perceptual data (Cox, 1946; Jaynes, 2003). Von Neumann and Morgenstern (1944) contributed utility theory as the formalized ideal for rational agents and their decision making. They presented an axiomatic formalization of preferences and derived the principle of maximum expected utility (MEU). This is computed for each action as the average utility of the action under consideration with respect to the probability of states of the world. The combinatorial complexity of real-world decision-making led to the heuristic models of bounded rationality by Herbert Simon in the late 1950s. Probabilistic and MEU decision-making methods were held back by assumptions that they are too inflexible, simplistic, and intractable. In the late 1980s, these assumptions got refuted, the methods proved widely useful, and they gained a large following in research, including AI. Attempts to mechanize probability and MEU processes led to thinking about the role of related representations and inference strategies in human cognition, which was followed by formulations of network-based representations, such as Bayesian networks, also called probabilistic graphical models (PGMs). They used belief-updating procedures with parallel and distributed computation. The ongoing research direction of PGM focuses on hierarchical models for capturing shared structures across data sets and active learning to guide the collection of data and probabilistic programming tools to specify context-sensitive models via compact high-level programs. The most important challenge is to move from classical utility-maximizing rational agents to a theory of resource-bounded systems in the complex and uncertain real world.

Especially important groundwork for computational rationality was laid by the conception of provably bounded-optimal agents by Russell & Subramanian (1995). They demonstrated that boundedness and optimality are not opposites at all. On the contrary, the two concepts need to go together if one wants to even approximate what “doing the right thing” in the real world could look like. Russell and Subramanian define how rationality is to be understood in contrast to the theoretical foundations that are centered around the unsatisfiable requirements of perfect rationality in intelligent systems. To close the historically already big gap between AI in theory and in practice, the authors suggested the property of bounded optimality (BO) in 1995.

Bounded optimality is defined as the solution to a constrained optimization problem presented to an agent by its architecture and the real-time task environment. Additionally, the authors propose a weaker property, which they call asymptotic bounded optimality (ABO), that generalizes the notion of optimality from classical complexity theory. They differentiate definitions and requirements of perfect rationality, calculative rationality, metalevel rationality, and bounded optimality:

- The classical notion of **perfect rationality** talks about an agent that maximizes its expected utility, given its acquired information. These agents cannot exist for non-trivial environments, as action selection requires computation and computation takes time.
- **Calculative rationality** captures the kind of agent that computes the decision to which it *should* have come as it *started* to compute it. This "in principle" capacity to do the right thing even for a chess program may take up to ten times longer and is therefore not interesting for any practical application. Russell and Subramanian explain that it is not clear that AI would have focused on the quest for calculative rationality had it not been started in the halcyon days before formal intractability results were discovered.
- **Metalevel rationality** optimizes over the object-level computations to be performed in the service of selecting actions. The major problem with this conceptualization is that the metalevel decision problem is often more difficult than the object-level problem. As a result, only approximations of metalevel rationality have shown to be useful (e.g., with metalevel policies that limit lookahead in chess programs).

- A **bounded optimal agent** behaves as well as possible given its computational resources in episodic real-time environments in which the utility of an action depends on the time at which it is executed.
- Analogical to classical complexity and computability theory where "absolute efficiency is the aim, asymptotic efficiency is the game", bounded optimality might only be useful as a philosophical landmark. The weaker form of **asymptotic bounded optimality** however can be used as a theoretical tool to actually make tractable the problems that universal real-time systems are facing. Russell and Subramanian (1995) make the point that one of the most interesting and worthwhile directions to take this research would be to question how this applies to learning agents that adapt to a shorter-term horizon and rewrite themselves as former adaptation becomes obsolete.

This imposes the question of meta-reasoning; the ability to make expected utility decisions about the ideal balance between effort or delay, and the quality of actions taken in the world. Rational meta reasoning has been explored in multiple problem areas, including guiding computation in probabilistic inference and decision-making and handling proactive inference in light of incoming streams of problems. The value of these questions is underscored by the complexity of probabilistic inference in the Bayesian network. This has been shown to be in the non-deterministic polynomial-time (NP)-hard complexity class and highlights the importance of developing approximations, e.g., by Monte Carlo simulations, bounding methods, and methods that decompose problems into simpler sets of subproblems. Analogous ideas have come to be increasingly important to how cognitive scientists and neuroscientists think about intelligence in human minds and brains, often being explicitly influenced by AI researchers and sometimes influencing them back.

Computational rationality makes the theoretical requirements and practical implications of how a learning agent embedded in a real-time environment, where computation is intrinsically costly in effort and externally in opportunity costs, can reach its goals. Gershman et al. (2015) argue that through this notion, the fields of artificial intelligence (AI), cognitive science, and neuroscience are reconverging on a shared view of the computational foundations of intelligence. They also synthesize insights from the two decades of scientific developments since the idea of asymptotic bounded optimality (Russell & Subramanian, 1995).

5.3.1 Rational decisions under bounded computational resources

"I won't look too far ahead
It's too much for me to take
But break it down to this next breath, this next step
This next choice is one that I can make
So I'll walk through this night
Stumbling blindly toward the light
And do the next right thing
And, with it done, what comes then?
When it's clear that everything will never be the same again
Then I'll make the choice to hear that voice
And do the next right thing."

- The Next Right Thing, Disney's Frozen II Soundtrack

Rational meta reasoning is essential for a decision-theoretic understanding of bounded rationality. The goal of metalevel analyses is to allow computational systems to find the ideal balance between effort, delay, and quality of actions taken in a given environment.

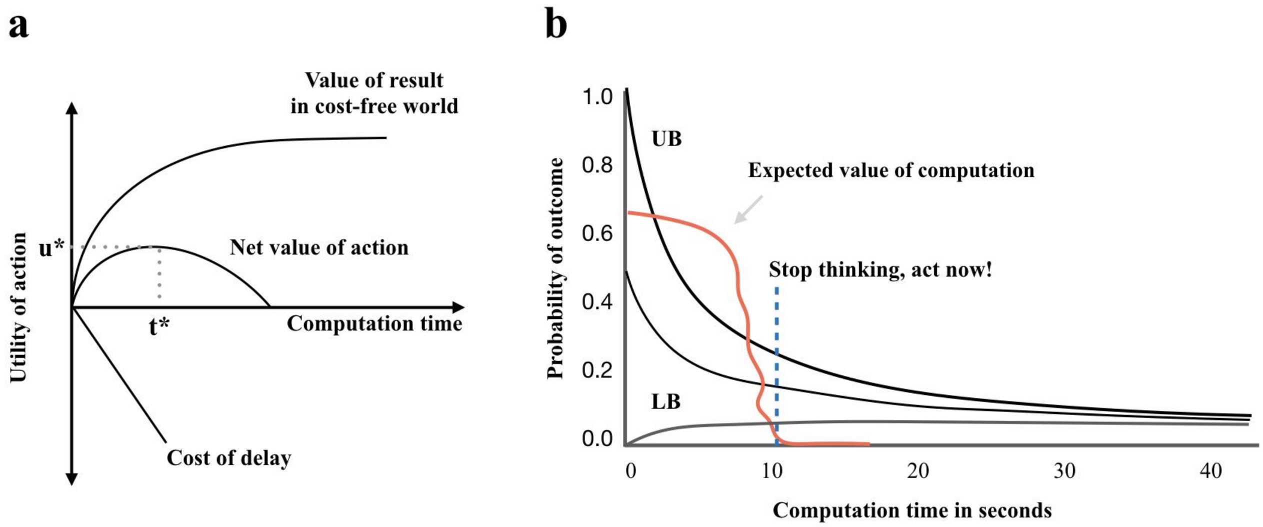

Figure 6. A visual depiction of trade-offs in computational rationality (Gershman et al., 2015) **a.** When costs increase linearly with delay, the portrayed value of action increases with decreasing marginal return. This results in an optimal stopping time, at which additional precision entails a net loss in action value. In the real world, the economics of cost and benefits of computation are uncertain on all fronts. **b.** The second part of the depiction shows a trace of rational meta reasoning in a time-critical medical setting. The bounding algorithm continues to tighten the upper and lower bound (UB & LB) on an important variable representing a patient's physiology. When a continually computed measure of the value of additional computation (red line) goes to zero, the base-level model is set off to make a recommendation for immediate action by taking the current best-inferred base level action ("stop thinking and act now!).

What then can computational rationality tell us about rational decisions under bounded computational resources embedded in a real-time environment? Perceiving, judging, and making decisions involve different computational costs that can be measured in different ways, from the loss in utility by delaying action in time-critical "online" reasoning to measures of effort applied or a number of inferences used among multiple inferential components (Gershman et al., 2015). The relation between value and cost of inference at different levels of accuracy is illustrated in Figure 6. The illustration shows, how, in a cost-free world, the value of additional effort applied and higher-utility action determined, only increases. When factoring in the opportunity cost of delay of action, the net value of taking starting to act quickly reaches a peak, an optimal time to stop thinking, after which the value of additional accuracy and effort falls off.

The trade-offs in and between different systems can be and often are uncertain themselves. The resulting worst-case complexity of the polynomial-time NP-hard complexity class of the problems to be solved highlights the importance of developing approximations that exploit the structure of real-world problems. The computational trade-offs are calculated differently depending on the decision-making system, such as, in the case of fast and inflexible model-free values in a look-up table or slower more flexible model-based forward search. AI systems, such as AlphaGo, use complex value function approximation architectures, such as deep convolutional nets for model-free control, and sophisticated forward search strategies, such as Monte Carlo Tree Search, for model-based control. Next to real-time meta-reasoning, offline analyses can help gather and improve on policies for guiding real-time meta-reasoning and for enhancing real-time inference via such methods as precomputing and caching parts of inference problems into fast-response reflexes.

Computational rationality[3] offers a framework to potentially unify the study of intelligence in minds, brains, and machines, with three central insights: 1) intelligent agents ultimately aim to

---

[3] According to Lieder and Griffith (Lieder & Griffiths 2020, Griffiths et al. 2015, Lieder et al. 2012) variations of the resource-rationality principle are known by various names, including computational rationality (Lewis et al. 2014), algorithmic rationality (Halpern & Pass 2015), bounded rational agents (Vul et al. 2014), boundedly rational analysis (Icard 2014), the rational minimalist program (Nobandegani 2017), and the idea of rational models with limited processing capacity developed in economics (Caplin & Dean 2015; Fudenberg et al. 2018; Gabaix et al. 2006; C. A. Sims 2003; Woodford 2014).

form beliefs and plan actions for maximizing expected utility; 2) these utility calculations may be intractable for real-world problems but can be effectively approximated by algorithms that seek a more general expected utility that takes into account the costs of computation; 3) the algorithms can be optimally fitted to the organism's specific needs. This can be done offline using engineering or evolutionary design, or online through meta-reasoning mechanisms for selecting the best approximation strategy in a given situation.

5.3.2 Resource-Rationality of Heuristics

Based on computational rationality one can undertake a resource-rational analysis. This analysis asks whether, given the constraints of costly computation and opportunity cost, a judgment was optimal. As a framework, resource rationality demonstrates the naiveté of a predefined set of prescriptions centered on more effortful thinking, more complex analysis, and mindware to increase rationality for an agent embedded in the real world. In short, thinking longer cannot always be more rational, as the agent has to optimize a speed-accuracy trade-off.

As elaborated above, many human decisions and judgments are said to systematically violate the laws of logic and probability theory (Tversky & Kahneman, 1974). These violations are known as cognitive biases. Cognitive biases have been interpreted as evidence that people do not reason by the rules of logic and probability theory but by simple, yet error-prone heuristics. Whether or not these findings prove that humans are irrational has been debated for decades (Stanovich, 2009).

Some view heuristics and the apparent biases that follow from them as a sign of irrationality. Reconceptualizing rationality to include its constraints by opportunity costs in the environment and computational costs suggests that some heuristics and the sometimes implied biases have to be understood as resource-rational strategies that cannot easily (if at all) be improved (Lieder et al., 2013; Griffiths et al., 2015; Lieder et al., 2014, Lieder & Griffiths, 2020). One so-called bias, stemming from the process of anchoring and adjusting, classically appears flawed when measured against the standards of logic and probability theory, that are typically used to assess rationality. However, the same effect could arise from an optimal trade-off between the costs caused by the error in judgment and the cost of the time it takes to reduce this error (Lieder et al., 2018; 2018). Combined with reasonable assumptions about human cognitive capacities and limitations, bounded optimality provides a more realistic normative standard for cognitive strategies and representations (Griffiths et al., 2015). An illustration of this lowered normative ceiling is depicted in Figure 7. The more realistic normative model also allows psychologists in the classical study of heuristics and biases to differentiate between actual systematic errors and biases that are only classified as such as an artifact of unrealistic standards, which can be traced back to Bayes-optimal heuristics[4].

[4] With resource-rationality in the framework of computational rationality, the normative bound of optimality is decreased to a bounded optimality notion. It has both implications for how high the ceiling of normative rationality can be, but also practical implications for prescriptions to increase rationality (Wallin, 2013).

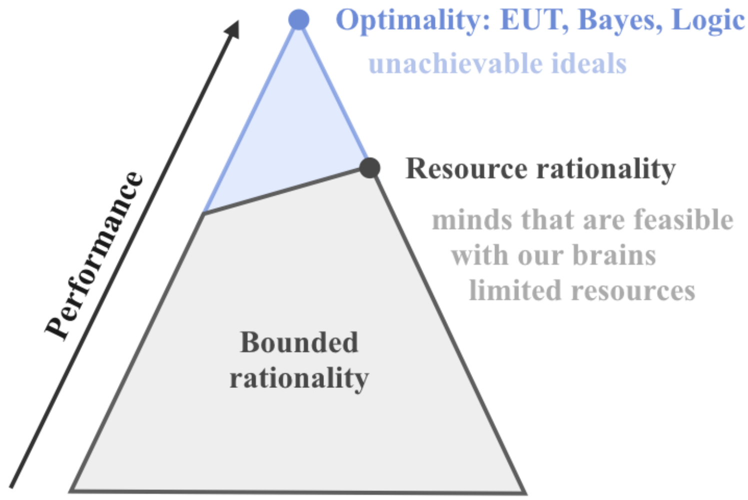

Figure 7: Resource rationality establishes a lowered normative standard for bounded rationality, by integrating knowledge about fundamental computational constraints into the unachievable ideal of Bayesian decision theory and logic (Lieder et al., 2018; Lieder & Griffiths, 2020)

Applying the resource-rational analysis to anchoring and its adjustments suggests that the number of adjustments people perform achieves a near-optimal trade-off between the costs of error and time. To test this assumption, a hypothesis can be formalized and tested by a computational model that chooses the number of adjustments to minimize the expected sum of time cost and error cost. Experiments have confirmed the predictions of resource-rational anchoring and adjustment (Lieder et al., 2012, 2013, 2017a, Shenhav et al. 2017). Most importantly, people appear to rationally adapt their number of adjustments to the summed cost. When errors were costly, people invested more time and were more accurate; their adjustments were larger, and their anchoring bias was smaller. The question of whether the adjustment changes gradually or in two distinct – fast and slow – modes (Evans, 2008; Kahneman, 2011) has yet to be examined and decided (Lieder et al., 2017). As a result, while people are biased according to the statistical meaning of the word, this bias might be consistent with how people ought to reason, following the ideas of computational rationality. Different so-called biases might not be identified as cognitive biases after applying resource-rational analysis. Many heuristics, while not intuitively Bayesian in their content, are Bayes optimal taking the cost of computation and opportunity cost into account.

5.4 From Simon's Scissors to Another Conflict

Computational rationality as a framework gives a sense of the limitations of a predefined set of prescriptions centered on more effortful thinking and complex analysis to increase rationality. When acting in an environment the agent has to approximately optimize a speed-accuracy trade-off in combination with adequate meta-reasoning. However, the speed-accuracy trade-off is, again, built on a strong and potentially wrong assumption, namely that thinking for a longer period of time or taking more aspects into account is always better. This is only half the rationality problem an organism or agent faces.

Herbert Simon (1990) introduced a conception of rationality that is shaped by both internal constraints and the task environment. He wrote: "human rational behavior (and the rational behavior of all physical symbol systems) is shaped by scissors whose two blades are the structure of task environments and the computational capabilities of the actor". Thinking in line with the scissors metaphor ("Simon's scissors"), the question of resource-rational analysis, computational rationality, and the underlying speed-accuracy trade-off puts the focus on only one blade of the scissors that make up rationality. The computational rationality approach to the axiomatic and ecological approaches encloses and approximates the two blades of Simon's scissors from two different sides, and moves them closer together, the more the sides integrate newly learned insights from the respective other. This Chapter started out by arguing that the prescription towards more effortful thinking did not take into account cost in computation and delay. However, there is yet another aspect missing from the analysis.

The underlying theme of computational rationality is the optimization of a speed-accuracy trade-off. It means that when performing a task, either speed or accuracy has to be sacrificed to achieve more of the other. It also assumes that quick responses lead to more errors and slow ones to fewer errors. It assumes that thinking or computing for longer would necessarily lead to a better outcome. However, taking a speed-accuracy trade-off into consideration,reveals the assumption that agents have an a priori sense of what the right computation for a given problem is. Two concepts help to understand why there is no optimal way of thinking, independently of the environment. These are the bias-variance dilemma and the no-free lunch theorems, described below. With their introduction, one can shift to a new notion of uncertainty that is a function of an dynamic organism-environment system, sometimes refered to as "fittedness" (Felin et al., 2017; Kozyreva & Hertwig, 2019, Vervaeke et al., 2012). The reconceptualization puts the question of the frame problem at the center of rationality and intelligence.

5.4.1 Bias-Variance Dilemma and Decomposition

Gigerenzer and Brighton (2015) coined the term “bias bias”. They argue that simple models, e.g., in marketing and finance, are sometimes on average more accurate, than more complex, sophisticated models. The same principle applies to human cognition. The strength of simple solutions is often misjudged because of a strong “bias bias” in behavioral economics and cognitive science.

To nourish this abstract explanation of trade-offs between multiple error sources with a more concrete understanding, let us think through a classic example from finance (Hertwig, 2019). The example bets the relative performance of “optimal” and “naive” diversification strategies against one another. DeMiguel et al. (2009) asked themselves how comparatively inefficient the 1/N portfolio strategy[5] is by evaluating out-of-sample performances. To do so, they evaluated 14 different models, including Markovitz’s (1952) Nobel prize-winning mean-variance model and the 1/N strategy, across seven empirical data sets. They concluded that across the datasets, no one model is consistently better than the 1/N rule in terms of Sharpe ratio, certainty-equivalent return, or turnover, which indicates that, out of sample, the gain from optimal diversification is more than only offset by estimation error. To outperform the 1/N benchmark the models would need 3000 months of data for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets. This means that different models attempting to optimize explicitly can outperform simple strategies like 1/N, but they do not do so consistently across out-of-sample data sets.

[5] The 1/n portfolio strategy is a portfolio allocation scheme that allocates an equal share of wealth to each of all (N) available investment possibilities.

Which model, or whether complex or simple strategies, work best, cannot be judged a priori, but depends on the task environment and the conditions (Hertwig, 2019; Nassar, 2020).

Another example to make the problem more concrete is locating the home of a serial criminal offender (Brighton, 2021). In this example, the serial offender is assumed to live at a single home location while committing a series of crimes. How can the location of the home be most accurately predicted? This well-studied problem in geographical crime profiling illustrates the relationship between variance and prediction error well. For this example, two geographical profiling models have been used: the centroid method and the probability surface model. The centroid method predicts a "center of gravity" between the crime locations, which corresponds to the mean x- and y- coordinate of the observed crimes. The second method uses an exponential decay function that estimates for each cell of the area the probability to contain the home address of the offender. The cell with the greatest estimated probability does indeed predict the true home location of the offender with close to zero error. In contrast, the centroid method predicts a point roughly 1.5 km northwest of the actual home location. The probability surface model is more accurate in this instance, but is it the better choice a priori for all cases? To answer this, the average performance of the two models over a range of serial criminal data has to be judged (e.g., Block & Bernasco, 2009; Leitner & Kent, 2009; Levine & Block, 2011).

The most important insight can be gained from deepening the uncertainty condition and comparing the relative performance of the methods. In the example, the number of data points of the crime locations was artificially decreased from initially 15 locations to only five. The outcome is that over various cases when fewer than 9 crimes are observed, the centroid method outperforms the probability surface model. The relative performance of the methods inverts when the conditions change. Again, this shows that there is no one technique that performs better a priori and independent of the circumstances. The central question changes from best method to fittedness to the question at hand.

At this point, it is important to stress the underdetermination problem (Quine, 1975; Laudan, 1990). It states that for any set of observations, there is an infinite number of logically consistent explanations. There is no ultimate way to confirm any given hypothesis against alternatives or to identify a best or true model in a platonic sense. Instead, multiple sources of errors face a trade-off.

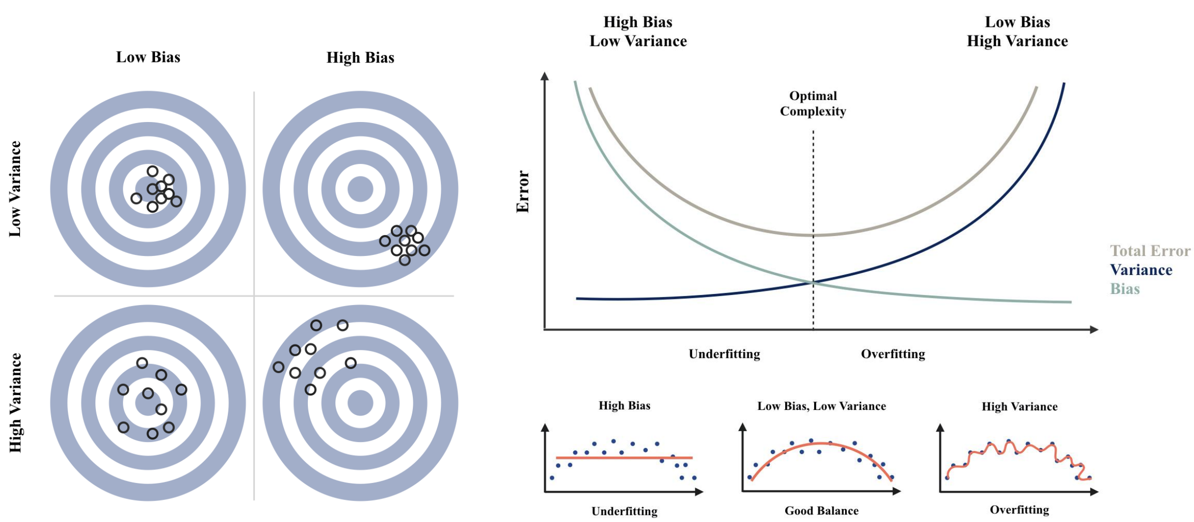

Figure 8. Diagramatic representation of bias-variance decomposition of error and relationship between bias, variance, and model complexity **a.** The total error of a model is decomposed as bias^2 + variance + noise. The diagram represents the relationship between variance and bias. **b.** illustrates how bias and variance relate to overfitting, underfitting, and optimal model complexity (Brighton & Gigerenzer, 2015; Gigerenzer, 2018; Gershman & Ölveczky, 2020)

In machine learning, finance, and decision theory, there are two sources of error that cannot both be minimized at the same time. One source is the bias error, the other the variance error. Techniques for reducing variance tend to increase bias, and techniques aimed at minimizing bias usually increase variance. This relationship is shown in Figure 8. The total error of a model is thus decomposed as $bias^2$ + variance + noise (van der Putten & van Someren, 2004; Brighton & Gigerenzer, 2015; Gigerenzer, 2018).

The bias error is based on erroneous assumptions in the learning algorithm. It points out how far away the mean of the prediction is from the points it tries to predict. The variance error is instead an error arising from sensitivity to small fluctuations in training sets. High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs. In machine learning and statistics, the bias and variance errors correspond to the problems of overfitting and underfitting. Overfitting means that a statistical model contains more parameters than justified by the data. Some of the residual variation (the noise) has been extracted into the parameters as if these points represented the underlying model structure. Underfitting occurs when a parameter is missing, for example, when applying a linear model to non-linear data.

The axiomatic approach to rationality has focused intensely on biases, without analyzing the problem an organism faces through a bias-variance-dilemma lens. Analogically speaking, the ecological approach points out that the axiomatic approach is systematically underfitting with its strong bias bias (Brighton & Gigerenzer, 2015). A concrete example for such a “bias bias” committed by the axiomatic approach to rationality is the choice to literally call the 1/N strategy to diversification the "naive diversification bias" (Read & Loewenstein, 1995; Benartzi & Thaler, 2001; Fernances, 2013; Kliger et al., 2014). The axiomatic solutions correspond too closely to a certain environment and are not robust independently of circumstances. How close a model matches, represents, or captures the underlying structure of a problem, the feature that leads to low potential for bias, is only one determinant of low prediction error. The other part of the question is how robust the predictions of the model are to different realizations of the problem, the feature that potentially leads to low variance is often even more important (Brighton, 2019).

5.4.2 The No-Free Lunch Theorems

The bias-variance dilemma originates in the fact that in statistical pattern recognition and machine learning, so-called "no free lunch theorems" apply. Averaging over all possible problems and environments, any two search algorithms or forecasting models will have the same predictive accuracy; the computational cost of finding a solution averaged over all problems in the class is the same for any solution method (Wolpert, 1996; Wolpert & Macready, 1997). Thus, for all possible performance measures, no search algorithm is better than another when its performance is averaged over all possible discrete functions. The no free lunch theorems show formally that there is no a priori algorithm or model that is better than every other one independently of the addressed problem set[6], clearly showing that there is no such thing as a general learning algorithm independent of its environment.[7]

The implicit assumptions in a model must match the characteristics of the problem at hand in order to yield accurate inferences and to prove adaptive. The bias and variance of different solutions stem from the interaction between properties of the model itself and properties of the task or environment (Bishop, 2006, Geman et al., 1992, Hastie et al., 2001, O'Sullivan, 1986, Van Der Putten and Van Someren, 2004). Decomposing error sources into the different components is crucial. Both error sources – fundamentally – cannot be minimized at the same time, because techniques for reducing variance tend to increase bias (reducing parameters), and techniques aiming at minimizing bias (increasing parameters) usually increase variance. This error decomposition results in new insights into the conflict between axiomatic and ecological rationality. While one focuses on bias reduction, the other prioritizes validity and fittedness to the environment. This means that the question of rationality is not simply the optimization of a speed-accuracy trade-off. According to Hertwig (2019), the human brain resolves the dilemma in the case of the typically sparse, poorly-characterized training sets provided through experience by adopting high-bias/low-variance heuristics. This reflects the fact that a zero-bias approach has

[6] Another angle of argumentation could follow Vapnik's necessary and sufficient conditions for consistency, which build on the point that specific constraintsthat tend to not be given are needed for generalization (Vapnik & Chervonenkis, 1981; Vapnik, 1998)

[7] The no-free-lunch theorem is also used by Schrittwieser (2020) as a point against the Singularitarian "Intelligence Explosion" argument (Bostrom, 1998; Kurzweil, 2005; Yudkowsky, 2004). Schrittwieser argues that in AI research "not only is there no free lunch, lunch gets more expensive every day". With this he means that not only does the theorem apply, in a practical sense, the additional intelligence acquired becomes more and more expensive in data, compute and research funding.

poor generalizability to new situations, and also unreasonably presumes precise knowledge of the true state of the world.

How to optimize the economical question of the meta-learner is undecidable a priori, as there is no optimal search strategy. The rationality question is fundamentally computationally irreducible (Wolfram, 2002). It is not possible to shortcut a program or to predetermine the future to see how the pattern unfolds. The resulting heuristics to tame the uncertainty are relatively simple, but produce better inferences in a wider variety of situations. The argument for simplicity in applied heuristics goes far beyond a mere reference to Occam's razor. The insightful perspective of ecological rationality is built on arguments of robustness to environmental uncertainty. In marketing, finance, but also in cognition, simple models can sometimes surprisingly lead to more accurate results than more sophisticated approaches, at least averaged over multiple potential instantiations of outcomes. The question then becomes, under which environmental circumstances does a particular strategy, whether it's simple or complex, work (Hertwig, 2019). At this point, metaphorically speaking, the rationality question jumps from one blade of Simon's Scissors to the other.

Taking computational and opportunity costs into account, many biases turn out to be a good approximation of an optimal solution to a speed-accuracy tradeoff. But computational or resource-rationality is not enough. Taking environmental and informational uncertainty as well as the no-free-lunch theorems into account, the speed-accuracy tradeoff does not hold. Simple solutions can be more robust to different environmental conditions that cannot be optimized for a priori.

5.4.3 Optimally overcoming Framing Effects

Looking back historically, one finds the framing effect and the underlying prospect theory as one of the core insights underlying behavioral economics and the axiomatic approach to rationality. Confronted with risky prospects, the possible outcomes and the probabilities of the outcomes have to be weighted. The same option can, linguistically or mentally, be framed or described in different ways (Tversky & Kahneman, 1979, 1981,1984). For example, the same possible outcomes of a gamble can be described as a certain probability of loss or as a certain probability of gains relative to the status quo.

The most well-known demonstration of a framing effect was presented by Tversky and Kahneman (1981), using the "Asian disease" thought experiment:

> *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 is as follows:*
>
> *Positive frame*
>
> *If Program A is adopted, 200 people will be saved.*
>
> *If Program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved.*

Negative frame

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.

All four described choices have the exact same expected value, when calculated by multiplying outcomes by their probability of occurrence, 200 lives saved and 400 lives lost. Program A and C are equivalent in associated probabilities of outcomes, as are programs B and D. While the options are equivalent, participants presented with the positive frame predominately chose Program A, the certain option (72 percent). The participants responding to the negative frame primarily chose Program D, or the risky option (78 percent) (Bloomsfield, 2006).

"Rational" invariance to framing effects requires that such changes in the description of outcomes, a shift of focus on "lives lost" or "lives saved", should not alter the preference order. People repeatedly demonstrate failures of invariance, which means they strongly react to the way the same risky choice is presented. The "Asian disease" thought experiment illustrates that a mere change in formulation changes a shift of preference from risk aversion to risk-seeking. This effect has been replicated extensively (e.g. Ruggeri et al., 2020).

One underlying argument of axiomatic rationality is that ignoring unnecessary context is crucial – but what constitutes unnecessary context? Imagine the "Asian disease" problem embedded in a practical situation. You are the director of a political mission and enter the board room; in the meeting that follows, you are presented with either the positive or the negative frame mentioned above. Viewing this through the lens of axiomatic and computational rationality, you are presented with a speed-accuracy trade-off. If you have enough time, you can use your System 2 reasoning to decouple from the current situation and actively reframe the options presented into a more neutral representation of the odds and values of the options you have to choose from. Asking “what are we even talking about", quantitatively, seems like a reasonable way to approach the question, which should accordingly lead to indifference to a choice between the positive and negative frame.

Being embedded into a social – and thus multi-agent, real-life – situation, the two formulations, however, can mean multiple other things (Gigerenzer, 1996). The formulation can be intentionally chosen by the presenter and entail further information about her experience with the subject at hand, including a reference to understand what the situation is about. Depending on the goals and mission of the organization you are working for, the choice might be relevant for you or not. Framing effects can also be exploited deliberately to manipulate the relative attractiveness of the options. Thaler (1980) noted that lobbyists for the credit card industry insisted that any price difference between cash and credit purchases be labeled a cash discount rather than a credit card surcharge. Depending on the intentions of the presenter, you might be intentionally misled to act in their interest, or they might want to give you a useful lead to fulfill your and your organizations’ goals. In one case taking the frame as an additional piece of information can be valuable, while in the other, an intentional decoupling of the information is the better choice.

In judgment and decision-making under uncertainty, no clear calculus of risky choices can be applied. In a multi-agent environment, the way information is presented can mean additional pieces of information or active deception that requires decoding of the symbolic representation. All of this takes place in a reference narrative against which background action is judged. The economists Kay and King (2020) pointed out that under radical uncertainty, the most fundamental question is not about inductive or deductive reasoning, but abductive reasoning. The central question is about understanding "what is going on here".

Dealing with the framing problem in a board meeting, the question then is what is the right thing to do? To which the answer is, as in a classical univeristy essay, “it depends”. What is the right thing to think or do? It radically depends on the context.

King and Kay (2020) mention a different example, represented in Figure 9, to understand the issue. In a triangle you read “A BIRD IN THE THE HAND”. In many experiments, when confronted with it, participants fail to spot the repetition of the article “the”. However, it is not clear whether that is an error. Having to deal with a text that comes with slight errors is something humans regularly have to solve. What the “right” or “rational” response is depends on the context. If you are at an optician and the text is hanging on the wall to test your eyesight, you are likely meant to read the words exactly how they are presented to you. If you are auditioning for a play, however, then what you see here is likely a typographical error in the text and meant to be ignored. What is relevant as an actor is probably your ability of emotional expression, so you better ignore the article. That which constitutes an error when reading the words in the triangle depends on the context. The so-called biases stemming from anchoring and adjustment and framing effects can be understood as computationally rational and therefore the optimal use of limited cognitive resources (Lieder et al., 2018; Lieder & Griffiths, 2020, Sharp et al., 2022). However, they indicate deeper statistical issues at the edge of the computational paradigm, the measurability of uncertainty, and the goal of optimality (Gershman et al., 2015; Brighton, 2018; Kozyreva & Hertwig, 2019; Brighton, 2020).

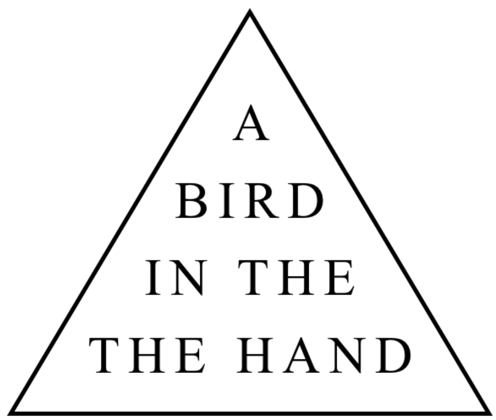

Figure 9. The triangle with the text "A BIRD IN THE THE HAND" is often referred to as an optical illusion, as the second "the" in the triangle is systematically overlooked. In the case of our work and King and Kay (2020) it is used to illustrate that what the correct way of reading the words in the triangle is not a priori defined. Accordingly, while one word is systematically not recognized, this is only to be considered an error in specific contexts.

5.5 Abductive Reasoning and Relevance Realization

Abductive reasoning and relevance realization allow a continuous transition between the statistical requirements of "small" and "large worlds". Brighton (2020) analysed the robustness of rationality principles against uncertainty conditions, differentiating between their use in small and large worlds. The former can be modeled with a decision matrix consisting of a mutually exclusive and exhaustive set of states of the world, consequences, and actions that lead from one to the other. Assuming a small world is a necessary condition (Savage, 1954) for applying Bayesian decision theory, although almost no real environment falls into this category (Binmore, 2009). In Brighton's differentiation, this means they are better understood as "large worlds". These are shaped by unmeasurable uncertainty (Knight, 1921; Kozyreva and Hertwig, 2019) and probabilistic ambiguity under the subjective interpretation of probability (e.g., Gilboa and Schmeidler, 1989; Schmeidler, 1992; Epstein, 1999). Brighton argues that it is more rational to regard generating distributions as non-existent than to view them as the foundation for formulating the rationality question or to consider their explicit application as the solution to the question.

As elaborated above computational or resource-rationality is not sufficient to deal with an uncertain environment because it assumes a priori knowledge of an optimal response. Given the no-free-lunch theorems, simple high-bias heuristics can be more predictive or adaptive to different potential instantiations of the structure of an unquantifiably uncertain environment (Gigerenzer & Todd, 1999). Figuratively speaking, the axiomatic and optimality-modeling approach and the ecological rationality tradition work towards each other from opposite ends of Simon's scissors. One is constrained by computational resources while presupposing the solution to the environment, the other primarily constrained by the structure of the environment. Both approaches are centered on different economical constraints and trade-offs. Now, the question is how can they be brought together?

Deductive and inductive reasoning have their role in understanding the world, once it is represented in closed symbol systems, but as one is confronted with larger worlds, the importance of abductive reasoning increases. Abductive reasoning – which may sound unfamiliar – means searching for the best explanation of what one observes (King & Kay, 2020). When events are essentially one-of-a-kind, which is often the case in the world of radical uncertainty, this kind of reasoning is indispensable. The core of abductive reasoning is finding out what is relevant to a situation, so-called *relevance realization*.

5.5.1 Relevance Realization

A major mistake of the General Problem Solver (GPS) framework of Simon, Shawn & Newell (1959) was to rely on the assumption that tasks form a well-defined category, or more broadly, that any problem could easily be turned into a well-defined or formal problem. For most real-world situations (e.g. taking notes during a lecture or taking a photo), every step of formalization is vague or even completely lacking: the initial state, the goal state, path constraints, and operators – all four components of the General Problem Solver. In large worlds, cognition has to deal with "ill-defined problems". The major skill in general problem solving and intelligence is to convert ill-defined problems into well-defined ones. This means making the search space more tractable by ruling out options that are not relevant (Vervaeke et al., 2012). When looking at how the dual-process theory and expert intuition are connected, expert chess masters perceptually reduce their search space while applying System 1 thinking[8]. Their

[8] The dual-process theory (Kahneman & Tversky, 1974; Stanovich & West, 2000; Kahneman, 2003; Kahneman, 2012) differentiates between Type 1 and Type 2 thinking, or System 1 and System 2 thinking,

perception and attention is directly shaped by their training, so they can already rule out moves that do not serve their goals by merely looking at the chessboard (Simon & Chase, 1973; Klein, 1998; Gawande, 2002; Chaiken & Trope, 1999; Gilbert, 2002; Sloman, 2002; Stanovich & West, 2002).

Related points about the primary importance of subjective perception for rationality are made by Felin, Krueger, and Koenderink (2017). Their argumentation also constitutes the target article of the Great Rationality Debate 2.0 (Chater et al., 2018). Felin et al. (2017) ground their work in the conception of the Umwelt by Uexküll (1921), which conceptualizes the semiotic world of an organism as grounded in that particular framing of the environment which includes all meaningful aspects of the world for a specific organism. They also stress the newfound importance for rationality of the ecological approach to perception by Gibson (1979), the patterns of animal behavior studied by Tinbergen (1951, 1963) and Lorenz (1955; Burkhardt, 2005), and Hoffmann et al.'s (2015) interface theory of perception.

The most important skill for achieving rationality in the "large world" of "ill-posed problems" is therefore the ability to focus on relevant information and the relevant structure of the information. It is cognitively transforming the large world into an addressable problem formulation or. a "small world". However, the smaller search space is only achievable by first considering the larger search space and dividing the relevant from the irrelevant.

referring to "fast or slow" modes of cognition. System 1 here means fast, automated, parallel, unconscious thinking processes, while System 2 means slow, effortful, sequential, controlled and conscious thinking processes.

To illustrate the depth of this challenge, Vervaeke et al. (2012) point towards a story about a robot written by Daniel Dennett:

A robot is trying to retrieve batteries as its source of food to then consume them. It comes upon a wagon with a battery on it, that could be retrieved by pulling the wagon. However, there is also a bomb located on the wagon, which would fall down as an unintended side effect. The designers of the robot, therefore, try to make sure they make the robot not only considers the intended effects of its actions but also unintended side effects, that would prevent the goal fulfillment. The robot, when put to the test, is not able to do anything anymore, as it just stands still and tries to go through all possible unintended side effects. The designers notice that only a small list of side effects is potentially relevant and design the robot to form a list of this sub list. This, however, only moves the problem of intractability from one problem to another. Once again the robot is immobilized, endlessly calculating the meta-list of relevant properties to consider (Dennett, 1984).

To determine what is relevant one has to determine first what is irrelevant. This entails, again, taking the whole “large world” search space into consideration, leading us back to the problem of combinatorial explosion. Relevance realization according to Vervaeke et al. (2012) seems paradoxical, like a circular definition that falls prey to recursive regress. The question to break through the theoretical circle of relevance realization is how to avoid combinatorial explosion and circular determination in focusing on what is relevant. The underlying paradox makes the problem a generalized version of the frame problem in artificial intelligence which has also been described as the “conundrum of the infinite regress of the Cartesian Theater” (Dennett, 1984).

5.5.2 Embodied Dynamicism and the Emerging Framework

"There is no means of testing which decision is better, because there is no basis for comparison. We live everything as it comes, without warning, like an actor going on cold. And what can life be worth if the first rehearsal for life is life itself? That is why life is always like a sketch. No, "sketch" is not quite a word, because a sketch is an outline of something."

— Milan Kundera, The Unbearable Lightness of Being

With their work, Verveake et al. (2012) critically built on work by Sperber and Wilson (1986, 1995), who proposed a theory of relevance, and Oaksford and Chater (1995), who deduced a theory of information gain on its basis. To prevent circularity in the definition of relevance realization, which overcomes precisely the problem introduced above, Verveake et al. (2012) identify three guidelines for theorists.

First, the conclusion must be that relevance cannot be captured as one static property of information. There can be no context-independent meta-theory of relevant information because relevance is context-sensitive. Instead of a theory of relevance, there needs to be a theory for a self-organizing mechanism for relevance realization. Second, the theory of the mechanism is neither representational nor syntactic, but economic. Third, instead of being instantiated by a completely general-purpose learning algorithm, it must involve competitors between multiple competing learning strategies.

To implement these three insights into a more sophisticated approach that offers open-ended context sensitivity, Vervaeke et al. (2012) holds one has to overcome script, schema, and stereotype theories and their "tunnel vision" (Arthur, 1994; Haugeland, 1985; Thompson, 2007; Varela et al., 1991) by studying how organisms and agents are embodied and embedded in the actual world. A much more sophisticated account of relevance realization is needed, which breaks through this threatening circle by dynamically integrating features such as frequency and invariance, rather than following a strict, context-insensitive rule.

This requires a paradigmatic shift from one perspective in the cognitive sciences to another. Namely, towards embodied dynamicism and the emerging framework. This more recent paradigm portrays cognition as a self-organizing dynamic system instead of a physical symbol system. Cognition and cognitive processes are no longer seen as outside the environment but considered to emerge from nonlinear and circular causality of continuous sensorimotor interactions between brain, body, and environment. Cognition is regarded to be an temporal phenomenon and needs a dynamic system perspective (Thompson, 2007). Relevance realization is then the underpinning of cognition as the exercise of skillful know-how in situated and embodied action, not reducible to prespecified problem-solving.

The argument of the impossibility of a theory of relevance realization by Verveake et al. (2012) is analogous to the impossibility of a theory of biological fitness. It is not possible to create a fixed set of properties that constitute fitness overall. Depending on the circumstances and organisms discussed, the fittest properties may be completely different. Instead of being a fixed set of traits, fitness is a dynamic product of the organism-environment-system and its fittedness. While a theory of properties constituting fitness is impossible, there can indeed be the theory of the mechanism, in this case of natural selection, that brings forth biological fitness in a continuous process. Accordingly, there cannot be a theory of relevance, but a theory of the cognitive mechanism for determining relevance. And, when locating relevance realization at the center of rationality, the same logic applies to rationality. Instead of aiming for a theory of rationality, the question becomes of the mechanism that realizes rationality. Similarly, Felin et al. (2014) argue that in the context of bounded rationality questions of tractability and the assumption of NP-completeness ato be problematic. No fixed fitness landscape is given and no solution pre-exists as the the NK approach to evolutionary fitness landscapes assumes (Kauffman & Levin, 1987; Levinthal, 1997). The frame problem can accordingly not be solved algorithmically at all, but require open-endedness (Felin et al., 2014)

Vervaeke et al. (2012) propose that relevance is never explicitly calculated by the brain at all, but that the phenomena that emerge as relevant, are a result of the brain's attempt to dynamically balance economic requirements. Cognitive agents face different interactional problems that imply corresponding internal economic processes and involve competition between opposing goals. The internal economical trade-offs and interactional problems from which relevance emerges, suggested by Vervaeke et al. (2012), are cognitive scope (CS), cognitive tempering (CT), and cognitive prioritization (CP), which all are mutually constraining within an internal economic arena. The economic constraints and interactional problems stemming from external interactional properties are applicability, projectability, and flexible gambling. Underneath each of the six points mentioned is a conflict and trade-off that has to be managed in a self-organizing, dynamical way.

Figure 10. This is a depiction of the higher order oppositional processes on economic constraints underlying relevance realization. These oppositional processes emerge emerges from multiple economic constraints and interactional problems: Internal economic properties constitute cognitive scope, cognitive tempering, and cognitive prioritization; External Interactional Properties constitute applicability, projectability, and flexible gambling. In the original publication the underlying internal and external properties are also visually represented (Vervaeke et al. 2012).

Vervaeke et al. (2012) analyze the oppositional processes on economic constraints that underlie relevance realization on a higher level of abstraction, which is illustrated in Figure 10. They locate on one side, efficiency and selection (time difference learning and exploiting), and on the other side, resiliency and variation (inhibition of return and exploring). While there is a selective constraint to be as efficient as possible, this efficiency can lead to a loss of latent pre-adaptive functions with long-term value. Consequently, there is also an opposing constraint to be resilient and handle environmental perturbations with redundancy and variation. These oppositional processes are analogous to the bias-variance dilemma, or the sides of the conflict between optimality modeling and ecological rationality. In a sense, Vervaeke et al. (2012) relocate the Great Rationality Debate 2.0 into cognition itself.

6. Conclusion

6.1 Relocating the Rationality Debate into Cognition itself

From this line of argumentation it follows that, while there is an ongoing debate between the approaches and solutions to the rationality question by the axiomatic and ecological approaches, the same conflict takes and has to take place in cognition to realize rationality. The conflict and trade-offs ultimately reflect a dynamic fight between different economic constraints and questions of efficiency and robustness in cognitive systems themselves.

The chasm between the two halves of Simon's scissors will not be resolved by ruling one half out as being theoretically fully reducible to the other side. This involves both, the side of computational rationality, optimality modeling and resource-rational analysis (or other frameworks building on an axiomatic a priori approach) and the side of ecological rationality (building on robust solutions that are adaptable to the environment). Both need to be studied and their context-dependent relevance has to be dynamically integrated, both in science as well as in cognition itself. Whether this is the same conclusion, overlapping with or unifiable with the dual-process model of cognition is yet to be studied, a question outside of the scope of this work. Jumping to conclusions too quickly and understanding this argumentation as the same as the dual-process model of cognition would likely be misguided (Kahneman, 2003). On the contrary, multiple other strains of perspectives and results could follow from this line of reasoning.

6.2 Paradigm shift to Non-Propositional and Non-Algorithmic Cognition

This thesis aims towards giving fundamentally new perspective on judgment and decision making under unmeasurable or radical uncertainty. It sheds light on how to bring axiomatic and ecological rationality closer together by viewing bounded rationality through Herbert Simon's scissors metaphor and the two resulting conflicts that have to be enactively balanced. Instead of externally differentiating between small worlds and large worlds or ill- and well-posed problems (Brighton, 2018; Kozyreva et al., 2019; Kozyreva & Hertwig, 2019; Knight, 1921; Marr, 1979; Savage, 1945), this work locates the essential part of rationality in the process of continuously transforming ill-posed problems (large worlds) into well-posed problems (small worlds) via relevance realization (Vervaeke et al., 2012; Vervaeke & Ferraro, 2013).

Savage (1945), the founder of Bayesian decision theory, wrote:

"[...] to cross one's bridges when one comes to them means to attack relatively simple problems of decision making by artificially confining attention to so small a world that the 'Look before you leap' principle can be applied there. I am unable to formulate criteria for selecting these small worlds and indeed believe that their selection may be a matter of judgment and experience about which it is impossible to enunciate complete and sharply defined general principles."

Bridging rationality and relevance realization leads to putting a focus on the question of how to select contextually relevant small worlds. The rationality question has to be brought conceptually closer to the question of how to solve the frame problem that is also faced in artificial intelligence (Dennett, 1984). This process of relevance realization is related to sense-making and autopoiesis (Maturana & Varela, 1980; Varela et al. 1991; Russell et al., 1993), deep conceptual changes (Thagard, 2012), a reconfiguration of one's salience landscape, problem transformation by insight (Klein, 2008, 2013, 2015; Viale, 2021; Felin & Koenderink, 2022), abductive reasoning (MacCarthy & Hayes, 1969; Kay & King, 2020), and wisdom (Vervaeke et al., 2012; Vervaeke & Ferraro, 2013).

Following this line of reasoning, rationality has to be reconceptualized as the process of ongoing open-ended systematic insight that transforms ill-posed statistical "large worlds" into well-posed "small worlds". As a result this enables a salience landscap that presents affordances for adaptive — instrumentally rational — action in the world (Savage, 1954; Simon, 1973; Marr, 1976; Vervaeke et al., 2012; Klein, 2013; Brighton, 2018; Bloekpol et al. 2018; Macchi & Bagassi, 2012; 2019; Felin & Koenderink, 2022).

6.3 The New Rationality landscape

This work, instead of arguing against axiomatic rationality, optimality, and risk modeling, aims to relocate the validity of these methods into the socio-cultural-ecological ontological thicket (Wimsatt, 2007; Thompson, 2007). This challenges the notion that "[...] the only viable formulation of perception, thinking, and action under uncertainty is statistical inference, and the normative way of statistical inference is Bayesian" (Edelman & Shabazi, 2011, p. 198). Instead, the methods are a culturally acquired psychotechnology that can be meaningfully applied to a reconceived rationality; either as a tool for insight and problem transformation or as a tool employed once a small world has been identified in which decoupled reasoning leads to results.

“It is a kit of cognitive tools that can attain particular goals in particular worlds,” writes Steven Pinker about rationality (2021).

Evan Thompson (2007) writes that, according to cognitive anthropologist Edwin Hutchins (1995), a central feature of the cognitivist view is a confusing transfer of a metaphor from human culture to individual psychology. The paradigm of cognitivism locates computation (e.g. computational rationality) inside the individual’s mind, instead of adequately identifying human computation as sociocultural activity. Consequently, it is impossible to understand that the properties of computation do not belong to the person but to the interpersonal and socio-cultural system which includes the environment. Understanding the probability calculus as constituting the laws of thought (Boole, 1854) represents a fallacy. Computation, a act historically understood as humans calculating, does not reflect the properties of the individual but of the system in which the individual resides. Therefore, cognitivism addresses cognition precisely backward because abstracting away from the situated and embedded individual does not leave us with the essence of individual cognition, but with the essentials of the sociocultural environment and system. Hutchins (1995) argues that the physical symbol-system architecture is not a model of individual cognition, but rather of the operation of a sociocultural system from which the human actor has been removed.

The proposal is to put axiomatic rationality and probability theory back into the operation of a sociocultural system and hereby bring into focus the processes of cognition and the human actor. This does not mean ignoring the robustly empirical insights about cognitive processes in human cognition from the research built on said axiomatic assumptions (Ruggeri et al. 2021). Many prescriptions, notions of how to move closer to normative rationality, still hold. Such as the probability theory needed to overcome a wrong belief in the law of small numbers (Tversky & Kahneman, 1971), which can be viewed as a tool for insight into mechanisms of the environment to systematically overcome delusion. It explains how prescriptions like more effortful thinking and active open-mindedness, meaning to proactively generate different explanations for the same perceptual input, can enable the recognition of a better frame (Stanovich, West, & Toplak 2016).

Exactly because these tools are part of the operation of the sociocultural system the human is embedded in, their knowledge and application are often a crucial adaptation for a higher fittedness of the organism-environment in a world which exhibits features of modernity. This line of argumentation adds a new meaning to the book title "Rationality in the modern world" by Stanovich (2009), namely moving closer to the normative ideals of axiomatic rationality as a context-dependent adaptation to the properties of modernity (Weber, 1922/2019; Pinker, 2021). But, if the axiomatic approach is not at the core of rationality, how else should it be viewed? Thompson (2007) writes that a physical symbol system is an elaborated and culture-dependent form of human action. Instead of being bounded by a brain, skull, or skin, it extends into the surrounding of the individual. Therefore it is embodied, requires perception and motor activity, and is also embedded in a sociocultural system of symbolic cognition and (psycho)technology. Clark and Chalmers (1998), and Wilson (1994) add that the environment directly plays a crucial and active role in cognitive processes; it is not a mere contingent and external setting.

This relocation of rationality is aligned with the proposed statistical requirements of ecological rationality by Brighton (2018) based on Breimann (2001). It acknowledges the problems an organism faces when taming unmeasurable uncertainty (Kozyreva & Hertwig, 2019; Kay & King, 2020). Rationality is non-propositional cognition and needs to take newer paradigms of cognitive science – embodied dynamicism and enactivism – into account. This point has already started to be acknowledged more regularly by others. Rolla (2019) talks about "situating rationality into radically enactive cognition" and Petracca (2021) uses the label 'embodied bounded rationality' to study how moderate embodiment can offer a new perspective on Simon's bounded rationality while, on the opposite side of the embodied spectrum, the label 'embodied rationality' is employed to explore how radical embodiment can more deeply transform the idea of what is rational. Gigerenzer (2021) introduces the concept of *embodied heuristics,* innate or learned rules of thumb that exploit evolved sensory and motor abilities. He provides a case study of the gaze heuristic, which solves coordination problems from intercepting prey to catching a flying ball. Like other scholars asking related questions this thesis understands these developments as steps in the right direction. There is large unchartered territory in thinking about the rationality question from a biological and embodied perspective (Roli et al., 2021; Rolla, 2019; Petracca, 2015; 2021; Gordon et al., 2021; Gigerenzer, 2021; Lehman et al., 2022; Fan et al., 2022; Mastrogiorgio et al., 2022).

6.4 Current Context and High Potential of the Future Rationality Debate

The transformation of large worlds into small worlds via relevance realization and insight is at the core of this line of argumentation. The proposed reconceptualization of the rationality question opens up a new set of questions and future research possibilities.

This moves the work of Gary Klein around naturalistic decision-making (Klein, 2008) that focuses strongly on problem transformations and insight, closer to the ongoing rationality debate between axiomatic and ecological rationality. How precisely they do relate will have to be established in future research.

While many underlying assumptions of behavioral economics and the heuristics and biases tradition break down with the suggested reconception of rationality, most of both their empirical cognitive findings and prescriptions of the tradition continue to hold. A special space is reserved for Stanovich's tripartite model of the mind, encompassing the reflective mind (Stanovich, 2009). The prescription of active open-mindedness to overcome the "what you see it all there is"-effect can equally well be grounded in generating insight to move from one frame to another (Stanovich et al., 2016; Kahneman, 2011). Connecting the empirical works and suggestions for rationality enhancements to these more solid foundations are an opportunity for future work.

One important foundation is the argumentation for the grounding of normativity not in optimality but in the self-maintenance of an autopoietic system. This explains, for example, why the phenomenon of losses weighing heavier than gains, according to prospect theory (Kahneman &

Tversky, 1979; Ruggeri et a., 2020), is not a descriptive deviation from rational choice but follows the normativity of utility being grounded in autopoiesis. While gaining additional resources is positive, any potential harm to furthering the organisms existence has to be prioritized for taking action. Similar arguments are brought forth by Taleb regarding convexity and antifragility of systems (2011; 2014; 2018) and Kauffman (1993).

Relevance realization often requires systematic insights into the nature of problems and a change in the perceived salience landscape. This highlights the importance of the first-person view to rationality and implies a reevaluation of the relationship between epistemic and instrumental rationality. Axiomatic rationality assumes that probabilistic beliefs have to reflect the real frequencies of the environment. This thesis points towards the importance of non-veridical strategies proposed in the Interface Theory of Perception by Hoffman et al. (2015) and in categorization (Hofstadter & Sander, 2011), as well as to the critical account of perceptual obviousness by Felin et al. (2017). Non-veridical strategies then raise questions about the functionality of attention and consciousness (Fazekas & Nanay, 2020; Baars, 1988). Additionally, these points have implications for the philosophy of science, where it is overdue to let go of the "view from nowhere" (Vervaeke, 2019) and the omniscient Laplacean demon. Instead a new impotence can be found in first-person accounts, subjectivity, and in the end "re-engineering philosophy" of science "for limited beings" (Wimsatt, 2007).

As Kay & King (2020) demonstrate, the immeasurability of uncertainty underlying the rationality debate (Kozyreva, 2019) and the resulting need for abductive reasoning is also crucial for economic theory and praxis. Accordingly, this work agrees with Kay and King (2020) that the concept of relevance realization can add an important understanding of the limitations of currently employed methods also to economics and gives a solid cognitive foundation to build its future. As the field of artificial intelligence was originally meant to be called computational rationality, the work of rethinking rationality in the cognitive sciences also has implications for current endeavors in AI, machine learning, and applied mathematics (Weinan, 2021). This is related to the argumentation of Roli, Jaeger, and Kauffman (2021), about fundamental limitations of potential AI systems. Of special interest will be the question of predetermined representations and the resulting limitations.

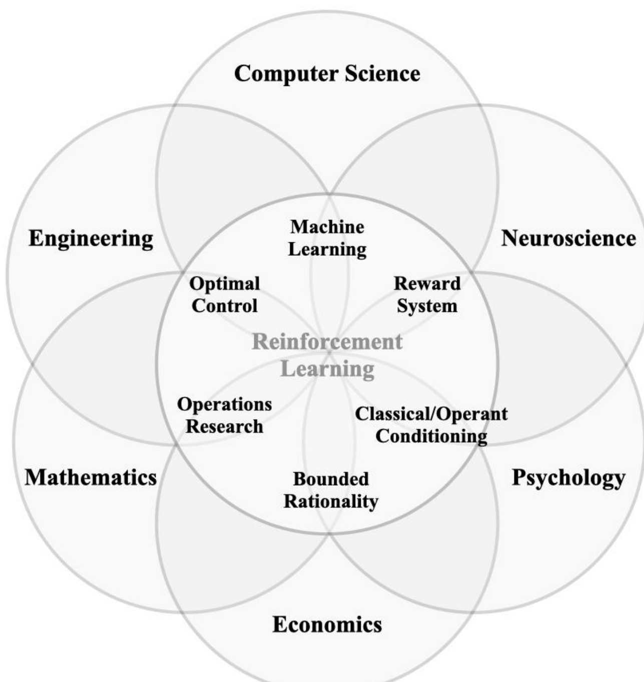

Figure 11. Reinforcement learning studies optimal decision-making in interaction with the environment. The Venn diagram depicts how this question, ultimately the same as the rationality question asked in this thesis, is asked as interdisciplinary dialogue in multiple disciplines (Silver, 2015)

New work and subsequent scissions in the intersection of topics around rationality and cognition (Felin et al., 2017; Chater et al., 2018; Gigerenzer, 2019; Brighton, 2018, 2020) support a narrative that understands the rationality debate not as a singular event of the past, but an ongoing process. It has to be crucially understood as part of even larger research questions and endeavours, which opens up further surface for exchange between fields and opportunity for insights.

The rationality question interfaces with many disciplines. These respective fields have experienced a surge in activity, continuously gathered new evidence, refined their positions, and formulated new theories. Future work to be done in cognitive science, additionally to merging the already explicitly communicated lines of arguments, is also going to be the integraton, unification and sense-making of vast spaces of insight by other disciplines through the lens of the rationality question. In a sense we are always just starting out.

Here are some examples of advancements in fields the rationality question interfaces with, to illustrate the larger landscape this discussion is part of:

- Bayesian epistemology constitutes an important subtopic of the normative theories of rational choice aiming for bayesian rationality. Works in bayesian epistemology and philosophy has surged to an degree that some call the last decade the bayesian era in philosophy of science (Oaksford & Chater, 2007; Colombo & Hartmann, 2017; Sprenger & Hartmann, 2019; Peden, 2020; Titelbaum, 2013, 2022, 2022). In recent years both analytic philosophy and cognitive psychology in general have made dramatic advances in understanding rationality (Broome, 2019; Knauff & Spohn, 2021; Hertwig et al., 2019; Viale, 2021).

- At the edge of mathematical finance, complexity science and economics, a questions related to rationality is that of the efficiency-robustness trade-offs in collective systems and institutions (Sunstein, 2018; Kahneman et al., 2021; Brighton & Gigerenzer, 2012). The robustness of simple heuristics could be part of the equation (Taleb, 2012; 2017, Devins et al., 2015). Understanding and taking into account the limitations of the current approach to risk modelling under the radical uncertainty of large world is indispensable (Kay & King, 2020).
- As also pointed at in the introduction, artificial intelligence in general aims to understand and create rational agents from first principles. The recent astonishing advancements in artificial intelligence, machine learning, computer science, adaptive computation, and applied mathematics accordingly offer many new insights into the rationality question (Russell & Norvig, 2020). Additionally Google's artificial intelligence program AlphaGo, in its match with Go player Lee Sedol, also generated a piece of cultural knowledge, move 37, that can now be taught to humans to make them more rational (Silver et al., 2017; 2018). This bidirectional exchange offers countless opportunities for future interdisciplinary exchange of ideas.
- Both artificial intelligence and the interdisciplinary reinforcement learning subfield examine the question of optimal decision-making in interaction with the environment. Reinforcement learning shows clearly how the rationality question borders to analogous questions in other disciplines, as shown the Venn-Diagram that constitutes Figure 11:

Bounded rationality in economics and psychology, operations research in mathematics, economics, and engineering, optimal control in mathematics, engineering, finance, and computer science, machine learning in engineering, computer science, and neuroscience, and reward systems in computer science, neuroscience, and psychology (Silver, 2015; Sutton & Barto, 2018).

- Silver et al. (2021) argue “that intelligence, and its associated abilities, can be understood as subserving the maximisation of reward”, which would imply that rationality could be fully understood through the (deep) reinforcement learning framework (Sutton & Barto, 2018). Silver et al. (2021) further write that “reward is enough to drive behaviour that exhibits abilities studied in natural and artificial intelligence, including knowledge, learning, perception, social intelligence, language, generalisation and imitation.”. The first principle approach to this research area has grown exponentially both in theory and in the creation of reinforcement learning agents. Additionally advancements in the neurobiology of (deep) reinforcement learning were achieved (Gershman & Ölveczky, 2020). The influences also come full circle, so has for example the dual-process theory from cognitive psychology, which was built on axiomatic rationality, been imported back into improving reinforcement learning systems (Kahneman, 2003, Botvinick et al., 2019). More interdisciplinary conversation is asked for.
- The ecological rationality approach to bounded rationality among many other advancements, invented boosting, researched the question of deliberate ignorance, the ecological rationality of the wisdom of crowds, attention allocation and stopping rules for search and learning strategies, the ecological rational choice depending on the structure

of the environment, ecological rational heuristics in strategic uncertainty, (Pleskac & Hertwig, 2014; Hertwig et al., 2019; Spiliopoulos & Hertwig, 2020; Hertwig et al., 2021, Hertwig & Engel, 2021; Viale, 2021)

The most important answers related to the rationality question continue to lie in the future and in retrospect what has happened up to now will be seen as prologue. We can define and ask the rationality question for multiple reasons, whether it is to ultimately understand our minds and therefore ourselves more deeply, to find ways to increase individual, collective, and institutional human rationality, or to build and safeguard intelligence in machines. Asking the rationality question continues to be theoretically intriguing and practically relevant.

7. Appendix

7.1 Abstract

The research endeavor concerning rationality as judgment and decision-making under uncertainty, or intelligent behavior in brains, minds, and machines, has repeatedly been a melting pot of interdisciplinary debates. At its recent peak, the axiomatic or optimality modeling approach to rationality and ecological rationality opposed each other.

This thesis argues that these two positions can be meaningfully triangulated through the lens of a Wimsattian robustness analysis derived from scientific perspectivism. The underlying goal is to divide the disagreements that merely constitute artefactual byproducts of the fundamental assumptions made from insights into the phenomenon of rationality from different perspectives.

Axiomatic and ecological rationality can be moved closer together by integrating computational rationality. Traditionally, it is held that taking computational constraints of cognition into account; rational agents face a speed-accuracy trade-off. Heuristics often meet the resulting lowered normative ceiling of resource-rational optimality.

The conception of this trade-off lacks a crucial element: Herbert Simon's scissors analogy indicates that bounded rationality is limited both by internal cognitive constraints and the task environment. Examining heuristics through the bias-variance dilemma any organism faces in an unknown territory adds the efficiency-robustness trade-off. In statistical "large worlds" or ill-posed problems, these two conflicts cannot be optimized a priori but must be negotiated in an emerging bottom-up manner by continuously resolving the frame problem. This process of overcoming the frame problem is called 'relevance realization.' It rests on problem transformation, sense-making, abductive reasoning, or insight. Therefore, the central question of rationality changes from a priori optimality to an ongoing optimal fittedness of an organism-environment system. This argument demonstrates the limitations of axiomatic rationality and reconceives it as a sociocultural tool for "small worlds" once the statistical requirements of the "large world" have been transformed by relevance realization. These cognitive findings have implications for categorization and perception, philosophy of science, economic theory, machine learning, and applied mathematics.

7.2 Kurzfassung

Die Forschungsarbeit zum Thema Rationalität als Urteils- und Entscheidungsfindung unter Unsicherheit — beziehungsweise Intelligenz im Gehirn, der Kognition und Maschine — wurde wiederholt ein Schmelztiegel interdisziplinärer Debatten. Auf dem jüngsten Höhepunkt standen sich axiomatische Rationalität und ökologische Rationalität gegenüber.

Diese Thesis argumentiert, dass diese beiden Positionen sinnvoll durch die Linse einer Wimsattschen Robustheitsanalyse trianguliert werden können, die vom wissenschaftlichem Perspektivismus abgeleitet wird. Das zugrunde liegende Ziel ist es, Meinungsverschiedenheiten, die lediglich artefaktische Nebenprodukte der Grundannahmen sind, von tatsächilchen Erkenntnissen über das Phänomen der Rationalität aus verschiedenen Perspektiven, zu differenzieren.

Axiomatische und ökologische Rationalität können durch die Integration von "Computational Rationality" weiter zusammengeführt werden. Bei der Berücksichtigung von rechnerischen Einschränkungen der Kognition wird traditionell davon ausgegangen, dass rationale Agenten einen Kompromiss zwischen Geschwindigkeit und Genauigkeit optimieren müssen. Heuristiken erreichen oft die daraus resultierende niedrigere normative Obergrenze der ressourcenrationalen Optimalität.

Der Konzeption dieses Kompromisses fehlt jedoch ein entscheidendes Element. Die Scherenanalogie von Herbert Simon zeigt, dass begrenzte Rationalität sowohl durch interne kognitive Einschränkungen als auch durch die Umwelt einbeschränkt ist. Die Untersuchung von Heuristiken durch das Bias-Varianz-Dilemma, mit dem jeder Organismus in einer ungewissen Lebenswelt konfrontiert ist, fügt den Kompromiss zwischen Effizienz und Robustheit hinzu. In statistischen "großen Welten" oder schlecht dargestellten Problemen können diese beiden Konflikte nicht a priori optimiert werden, sondern müssen durch kontinuierliche Lösung des Rahmenproblems verhandelt werden. Dieser Prozess der Überwindung des Rahmenproblems wird als „Relevanzrealisierung" bezeichnet. Relevanzrealisierung beruht auf Problemtransformation, Sinnstiftung, abduktivem Denken oder systematischer Einsicht. Damit verändert sich die zentrale Rationalitätsfrage von einer apriorischen Optimalität zu einer fortwährenden unbeschränkten Anpassung eines Organismus-Umwelt-Systems. Dieses Argument zeigt die Grenzen der axiomatischen Rationalität auf und begreift sie als soziokulturelles Werkzeug für statistische "kleine Welten". Diese sind gegeben sobald die statistischen Anforderungen der "großen Welt" durch Relevanzrealisierung transformiert wurden. Diese kognitiven Erkenntnisse haben Auswirkungen auf Kategorisierung und Wahrnehmung, Wissenschaftstheorie, Ökonomie, maschinelles Lernen und angewandte Mathematik.

References

Abel, D. (2019). *Concepts in Bounded Rationality: Perspectives from Reinforcement Learning*. 148.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*(2), 227–257. https://doi.org/10.1016/S0160-2896(96)90016-1

Al-Najjar, N. I., & de Castro, L. (2011). Subjective Probability. In J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, & J. C. Smith, *Wiley Encyclopedia of Operations Research and Management Science* (p. eorms0849). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470400531.eorms0849

Anderson, J. R. (1991). Is human cognition adaptive? Behavioral and Brain Sciences, 48.

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence? *Decision*, *3*(1), 20–39. https://doi.org/10.1037/dec0000043

Arrow, K. J. (1982). Risk Perception in Psychology and Economics. *Economic Inquiry*, *20*(1), 1–9. https://doi.org/10.1111/j.1465-7295.1982.tb01138.x

Arthur, W. B. (1994). Inductive Reasoning and Bounded Rationality. *The American Economic Review*, *84*(2,), 406–411.

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. In *Progress in Brain Research* (Vol. 150, pp. 45–53). Elsevier. https://doi.org/10.1016/S0079-6123(05)50004-9

Baddeley, M. (2017). *Behavioural Economics: A Very Short Introduction*. Oxford University Press. https://doi.org/10.1093/actrade/9780198754992.001.0001

Barett, L., & Salovey, P. (2002). *The wisdom in feeling*.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The Adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press.

Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty*, *14*(3), 301–309.

Baron, J. (2006). *Thinking and Deciding*. 600.

Baron, J. (2008). *Thinking and deciding* (4th ed). Cambridge University Press.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, *113*(3), 628–647. https://doi.org/10.1037/0033-295X.113.3.628

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual Differences in Working Memory Capacity and Dual-Process Theories of the Mind. *Psychological Bulletin*, *130*(4), 553–573. https://doi.org/10.1037/0033-2909.130.4.553

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of Self-Control. *Current Directions in Psychological Science*, *16*(6), 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x

Bayes. (1900). An Essay Towards Solving a Problem in the Doctrine of Chances. *Biometrika*, *45*(3–4), 296–315. https://doi.org/10.1093/biomet/45.3-4.296

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1–3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Benartzi, S., & Thaler, R. H. (2001). Naive Diversification Strategies in Defined Contribution Saving Plans. *The American Economic Review*, *91*(1), 79–98.

Bernasco, W., & Block, R. (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology*, *47*(1), 93–130. https://doi.org/10.1111/j.1745-9125.2009.00140.x

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23–36. JSTOR. https://doi.org/10.2307/1909829

Binmore, K. G. (2009). *Rational decisions*. Princeton University Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Bishop, M., & Trout, J. D. (2005). The Pathologies of Standard Analytic Epistemology. *Noûs*, *39*(4), 696–714.

Bloomfield, A. N., Sager, J. A., Bartels, D. M., & Medin, D. L. (2006). Caring about framing effects. *Mind & Society*, *5*(2), 123–138. https://doi.org/10.1007/s11299-006-0013-3

Bogers, M., Dahlander, L., Felin, T., Gorbatai, A. D., Reitzig, M., Sims, J., & Sundararajan, A. (2015). What is New About "Openness"? Implications for Value Creation and Value Capture. *Academy of Management Proceedings*, *2015*(1), 13776. https://doi.org/10.5465/ambpp.2015.13776symposium

Bostrom, N. (1998). How long before superintelligence? *International Journal of Futures Studies*, *2*.

Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences*, *23*(5), 408–422. https://doi.org/10.1016/j.tics.2019.02.006

Bourgeois-Gironde, S., & Giraud, R. (2009). Framing effects as violations of extensionality. *Theory and Decision*, *67*(4), 385–404. https://doi.org/10.1007/s11238-009-9133-7

Brady, M. E. (2019). What Keynes Stated in 1937, As He Had Stated in 1921 and 1936, Was That 'We Simply Do Not Know (About the Distant or Far Future)': The Post Keynesian (Robinson, Shackle, Weintraub, Davidson) Claim That Keynes Said Only That "We Simply Do Not Know" Period is a Total Fabrication. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3401412

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*(2), 409–432. https://doi.org/10.1037/0033-295X.113.2.409

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–215.

Brighton, H. (2018). *Rationality without optimality: Bounded and ecological rationality from a Marrian perspective* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/m2sz5

Brighton, H. (2020). Statistical foundations of ecological rationality. *Economics*, *14*(1), 20200002. https://doi.org/10.5018/economics-ejournal.ja.2020-2

Brighton, H., & Gigerenzer, G. (2015). The bias bias. *Journal of Business Research*, *68*(8), 1772–1784. https://doi.org/10.1016/j.jbusres.2015.01.061

Brighton, H., & Gigerenzer, G. (2012). Are rational actor models “rational” outside small worlds? In S. Okasha & K. Binmore (Eds.), *Evolution and Rationality* (1st ed., pp. 84–109). Cambridge University Press. https://doi.org/10.1017/CBO9780511792601.006

Broome, J. (2021). *Normativity, rationality, and reasoning: Selected essays* (First edition). Oxford University Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470

Burkhardt, R. W. (2008). Lorenz, Konrad Zacharias. In John Wiley & Sons, Ltd (Ed.), *ELS* (1st ed.). Wiley. https://doi.org/10.1002/9780470015902.a0002865

Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., & Griffiths, T. L. (2017). *A resource-rational analysis of human planning*. 7.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P., Lieder, F., & Griffiths, T. (2021). *Human planning as optimal information seeking* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/byaqd

Callebaut, W. (2012). Scientific perspectivism: A philosopher of science’s response to the challenge of big data biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 69–80. https://doi.org/10.1016/j.shpsc.2011.10.007

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Camerer, C., Loewenstein, G., & Rabin, M. (Eds.). (2004). *Advances in behavioral economics*. Russell Sage Foundation ; Princeton University Press.

Campitelli, G., & Gobet, F. (2005). The mind’s eye in blindfold chess. *European Journal of Cognitive Psychology*, *17*(1), 23–45. https://doi.org/10.1080/09541440340000349

Caplin, A., & Dean, M. (2015). Revealed Preference, Rational Inattention, and Costly Information Acquisition. *American Economic Review*, *105*(7), 2183–2203. https://doi.org/10.1257/aer.20140117

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199207077.001.0001

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Chaiken, S., & Trope, Y. (Eds.). (1999). Dual-process theories in social psychology. *Dual-Process Theories in Social Psychology.*, xiii, 657–xiii, 657.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, *2*(6), 206–214. https://doi.org/10.1016/S1364-6613(98)01179-6

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chater, N., Felin, T., Funder, D. C., Gigerenzer, G., Koenderink, J. J., Krueger, J. I., Noble, D., Nordli, S. A., Oaksford, M., Schwartz, B., Stanovich, K. E., & Todd, P. M. (2018). Mind, rationality, and cognition: An interdisciplinary debate. *Psychonomic Bulletin & Review*, *25*(2), 793–826. https://doi.org/10.3758/s13423-017-1333-5

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65. https://doi.org/10.1016/S1364-6613(98)01273-X

Chater, N., & Oaksford, M. (Eds.). (2008). *The Probabilistic Mind: Prospects for Bayesian cognitive science*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199216093.001.0001

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19. JSTOR.

Colander, D., Holt, R., & Rosser, B. (2004). The changing face of mainstream economics. *Review of Political Economy*, *16*(4), 485–499. https://doi.org/10.1080/0953825042000256702

Colombo, M., & Hartmann, S. (2017). Bayesian Cognitive Science, Unification, and Explanation. *The British Journal for the Philosophy of Science*, *68*(2), 451–484. https://doi.org/10.1093/bjps/axv036

Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature*, *34*(2), 669–700. JSTOR.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Cosmides, L. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*(1), 1–73. https://doi.org/10.1016/0010-0277(95)00664-8

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In *The adapted mind: Evolutionary psychology and the generation of culture.* (pp. 163–228). Oxford University Press.

Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, *14*(1), 1–13. https://doi.org/10.1119/1.1990764

Davis, J. B. (2003). *The theory of the individual in economics: Identity and value*. Routledge.

Davis, J. B. (2007). The turn in recent economics and return of orthodoxy. *Cambridge Journal of Economics*, *32*(3), 349–366. https://doi.org/10.1093/cje/bem048

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Dawes, R. M. (1999). *A message from psychologists to economists: Mere predictability doesn't matter like it should (without a good story appended to it)* ឲ. 12.

Dawes, R. M., Gilbert, D., Fiske, S., & Lindzey, G. (1998). *The handbook of social psychology*.

Dawkins, R. (2016). *The extended selfish gene* (40th anniversary edition). Oxford University Press.

De Neys, W. (Ed.). (2018). *Dual process theory 2.0* (1 Edition). Routledge, Taylor & Francis Group.

de Sousa, R. (2007). Truth, Authenticity, and Rationality. *Dialectica*, *61*(3), 323–345. https://doi.org/10.1111/j.1746-8361.2007.01104.x

Dekker, E., & Remic, B. (2019). Two types of ecological rationality: Or how to best combine psychology and economics. *Journal of Economic Methodology*, *26*(4), 291–306. https://doi.org/10.1080/1350178X.2018.1560486

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/ *N* Portfolio Strategy? *Review of Financial Studies*, *22*(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

Dennett, D. C. (1984). *Cognitive wheels: The frame problem of AI*. 13.

Dennett, D. C. (1996). *Kinds of minds: Toward an understanding of consciousness* (1st ed). Basic Books.

Devins, C., Koppl, R., Kauffman, S., & Felin, T. (2015). Against Design. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2570030

Diener, E., & Biswas-Diener, R. (2017). *The Replication Crisis in Psychology*.

Dugas, M. J., Gagnon, F., Ladouceur, R., & Freeston, M. H. (1998). Generalized anxiety disorder: A preliminary test of a conceptual model. *Behaviour Research and Therapy*, *36*(2), 215–226. https://doi.org/10.1016/S0005-7967(97)00070-3

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., Bishop, S., & Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General*, *137*(1), 131–148. https://doi.org/10.1037/0096-3445.137.1.131

Edelman, S., & Shahbazi, R. (2011). Survival in a world of probable objects: A fundamental reason for Bayesian enlightenment. *Behavioral and Brain Sciences*, *34*(4), 197–198. https://doi.org/10.1017/S0140525X11000252

Edwards, W. (n.d.). The Theory of Decision Making. 38.

Edwards, W., & Slovic, P. (1965). Seeking Information to Reduce the Risk of Decisions. *The American Journal of Psychology*, *78*(2), 188–197. JSTOR. https://doi.org/10.2307/1420490

Ellsberg, D. (1961). *Risk, Ambiguity, and the Savage Axioms*. 28.

Epstein, S. (2003). Cognitive-Experiential Self-Theory of Personality. In I. B. Weiner (Ed.), *Handbook of Psychology* (p. wei0507). John Wiley & Sons, Inc. https://doi.org/10.1002/0471264385.wei0507

Evans, J., & Frankish, K. (Eds.). (2009). *In two minds: Dual processes and beyond*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230167.001.0001

Evans, J. S. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Psychology Press.

Evans, J. S. B. T. (2017). *Thinking and reasoning: A very short introduction*. Oxford University Press.

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T., & Over, D. E. (2013). *Rationality and Reasoning* (0 ed.). Psychology Press. https://doi.org/10.4324/9780203027677

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Fan, L., Wang, G., Jiang, Y., Mandlekar, A., Yang, Y., Zhu, H., Tang, A., Huang, D.-A., Zhu, Y., & Anandkumar, A. (2022). *MineDojo: Building Open-Ended Embodied Agents with Internet-Scale Knowledge*. https://doi.org/10.48550/ARXIV.2206.08853

Fazekas, P., & Nanay, B. (2021). Attention Is Amplification, Not Selection. *The British Journal for the Philosophy of Science*, *72*(1), 299–324. https://doi.org/10.1093/bjps/axy065

Feiten, T. E. (2020). Mind After Uexküll: A Foray Into the Worlds of Ecological Psychologists and Enactivists. *Frontiers in Psychology*, *11*, 480. https://doi.org/10.3389/fpsyg.2020.00480

Felin, T., Kauffman, S., Koppl, R., & Longo, G. (2014). Economic Opportunity and Evolution: Beyond Landscapes and Bounded Rationality: Economic Opportunity and Evolution. *Strategic Entrepreneurship Journal*, *8*(4), 269–282. https://doi.org/10.1002/sej.1184

Felin, T., & Koenderink, J. (2021). Generative Rationality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3974977

Felin, T., & Koenderink, J. (2022). A Generative View of Rationality and Growing Awareness†. *Frontiers in Psychology*, *13*, 807261. https://doi.org/10.3389/fpsyg.2022.807261

Felin, T., Koenderink, J., & Krueger, J. I. (2017). Rationality, perception, and the all-seeing eye. *Psychonomic Bulletin & Review*, *24*(4), 1040–1059. https://doi.org/10.3758/s13423-016-1198-z

Felin, T., Koenderink, J., Krueger, J. I., Noble, D., & Ellis, G. F. R. (2021a). Data bias. *Genome Biology*, *22*(1), 59. https://doi.org/10.1186/s13059-021-02278-2

Felin, T., Koenderink, J., Krueger, J. I., Noble, D., & Ellis, G. F. R. (2021b). The data-hypothesis relationship. *Genome Biology*, *22*(1), 57. https://doi.org/10.1186/s13059-021-02276-4

Fernandes, D. (2013). The 1/N Rule revisited: Heterogeneity in the naïve diversification bias. *International Journal of Research in Marketing*, *30*(3), 310–313. https://doi.org/10.1016/j.ijresmar.2013.04.001

Fishburn, P. C. (1970). *Utility theory for decision making*. Wiley.

Flennerhag, S., Schroecker, Y., Zahavy, T., van Hasselt, H., Silver, D., & Singh, S. (2021). Bootstrapped Meta-Learning. *ArXiv:2109.04504 [Cs, Stat]*. http://arxiv.org/abs/2109.04504

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Foley, R. (1987). *The theory of epistemic rationality*. Harvard University Press.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Fudenberg, D., Strack, P., & Strzalecki, T. (2018). Speed, Accuracy, and the Optimal Timing of Choices. *American Economic Review*, *108*(12), 3651–3684. https://doi.org/10.1257/aer.20150742

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *THE AMERICAN ECONOMIC REVIEW*, *96*(4), 26.

Galesic, M., Olsson, H., & Rieskamp, J. (2012). Social Sampling Explains Apparent Biases in Judgments of Social Environments. *Psychological Science*, *23*. https://doi.org/10.1177/0956797612445313

Gawande, A. (2002). *Complications: A Surgeon's note on imperfect science*. Pan Books.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, *4*, 1–58. https://doi.org/10.1162/neco.1992.4.1.1

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. https://doi.org/10.1126/science.aac6076

Gershman, S. J., & Ölveczky, B. P. (2020). The neurobiology of deep reinforcement learning. *Current Biology*, *30*(11), R629–R632. https://doi.org/10.1016/j.cub.2020.04.021

Gibson, J. J. (n.d.). *The Ecological Approach to Visual Perception*. 347.

Gibson, J. J. (1974). *The perception of the visual world*. Greenwood Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Giere, R. N. (2006). *Scientific perspectivism*. University of Chicago Press.

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases." *European Review of Social Psychology*, *2*(1), 83–115. https://doi.org/10.1080/14792779143000033

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592–596. https://doi.org/10.1037/0033-295X.103.3.592

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. Oxford University Press.

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford University Press.

Gigerenzer, G. (2014). *Gut feelings: The intelligence of the unconscious*. Penguin Books. http://rbdigital.oneclickdigital.com

Gigerenzer, G. (2015). *Risk savvy: How to make good decisions*. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1124783

Gigerenzer, G. (2018). The Bias Bias in Behavioral Economics. *Review of Behavioral Economics*, *5*(3–4), 303–336. https://doi.org/10.1561/105.00000092

Gigerenzer, G. (2019). Axiomatic rationality and ecological rationality. *Synthese*, *198*(4), 3547–3564. https://doi.org/10.1007/s11229-019-02296-5

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, *1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1996a). *Reasoning the Fast and Frugal Way: Models of Bounded Rationality*. 20.

Gigerenzer, G., & Goldstein, D. G. (1996b). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011a). *Heuristics: The foundations of adaptive behavior*. Oxford University Press.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011b). *Heuristics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744282.001.0001

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox*. MIT Press.

Gigerenzer, G., & Sturm, T. (2012). How (far) can rationality be naturalized? *Synthese*, *187*(1,), 243–268.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In *Unintended thought.* (pp. 189–211). The Guilford Press.

Gilbert, D. T., & Ebert, J. E. J. (2002). Decisions and revisions: The affective forecasting of changeable outcomes. *Journal of Personality and Social Psychology*, *82*(4), 503–514. https://doi.org/10.1037/0022-3514.82.4.503

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*(2), 141–153. https://doi.org/10.1016/0304-4068(89)90018-9

Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037/0033-295X.109.1.75

Gordon, J., Maselli, A., Lancia, G. L., Thiery, T., Cisek, P., & Pezzulo, G. (2021). The road towards understanding embodied decisions. *Neuroscience & Biobehavioral Reviews*, *131*, 722–736. https://doi.org/10.1016/j.neubiorev.2021.09.034

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322. https://doi.org/10.1038/nn1014

Greindl, G. (2010). *Living in a natural world*. https://doi.org/10.25365/THESIS.9322

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian Models of Cognition. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (1st ed., pp. 59–100). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.006

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. https://doi.org/10.1111/tops.12142

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science*, *17*(9), 767–773. https://doi.org/10.1111/j.1467-9280.2006.01780.x

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*(7), 488–501. https://doi.org/10.1038/nrn3524

Hacking, I. (1990). *The Taming of Chance* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511819766

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

Halpern, J. Y., & Pass, R. (2014). Algorithmic Rationality: Game Theory with Costly Computation. *ArXiv:1412.2993 [Cs]*. http://arxiv.org/abs/1412.2993

Hammond, K. R. (1996). *Human judgement and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* (pp. xi, 436). Oxford University Press.

Harman, G. (1995). Rationality. In E. E. Smith & D. N. Osherson (Eds.), *Invitation to Cognitive Science*. MIT Press.

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In *Variation in working memory.* (pp. 227–249). Oxford University Press.

Hasher, L., Lustig, C., & Zacks, R. (2008). Inhibitory Mechanisms and the Control of Attention. *Variation in Working Memory*, *19*. https://doi.org/10.1093/acprof:oso/9780195168648.003.0009

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making* (2nd ed). SAGE.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009a). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed). Springer.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009b). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed). Springer.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea* (Vol. 38, Issue 151). MIT Press.

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00150

Hertwig, R. (2012). The psychology and rationality of decisions from experience. *Synthese*, *187*(1), 269–292. https://doi.org/10.1007/s11229-011-0024-4

Hertwig, R., & Engel, C. (Eds.). (2020). *Deliberate ignorance: Choosing not to know*. The MIT Press.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hertwig, R., Leuker, C., Pachur, T., Spiliopoulos, L., & Pleskac, T. J. (2021). Studies in Ecological Rationality. *Topics in Cognitive Science*, tops.12567. https://doi.org/10.1111/tops.12567

Hertwig, R., Pleskac, T. J., & Pachur, T. (2019). *Taming uncertainty*. MIT Press.

Heukelom, F. (2010). Measurement and decision making at the University of Michigan in the 1950s and 1960s. *Journal of the History of the Behavioral Sciences*, *46*(2), 189–207. https://doi.org/10.1002/jhbs.20425

Heukelom, F., & Sent, E.-M. (2017). *Behavioral economics: From advising organizations to nudging individuals*. 5.

Hicks, K. L., Harrison, T. L., & Engle, R. W. (2015). Wonderlic, working memory capacity, and fluid intelligence. *Intelligence*, *50*, 186–195. https://doi.org/10.1016/j.intell.2015.03.005

Hilton, D. (2003). Psychology and the financial markets: Applications to understanding and remedying irrational decision-making. *The Psychology of Economic Decisions*, 273–297.

Hiser, J., & Koenigs, M. (2018). The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology. *Biological Psychiatry*, *83*(8), 638–647. https://doi.org/10.1016/j.biopsych.2017.10.030

Hoffman, D. D. (2009). The Interface Theory of Perception: Natural Selection Drives True Perception to Swift Extinction. In S. J. Dickinson, A. Leonardis, B. Schiele, & M. J. Tarr (Eds.), *Object Categorization* (pp. 148–166). Cambridge University Press. https://doi.org/10.1017/CBO9780511635465.009

Hoffman, D. D., Singh, M., & Prakash, C. (2015). The Interface Theory of Perception. *Psychonomic Bulletin & Review*, *22*(6), 1480–1506. https://doi.org/10.3758/s13423-015-0890-8

Hofmann, N. (2020). *On the history of the rationality debate*. 3.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and Self-Control From a Dual-Systems Perspective. *Perspectives on Psychological Science*, *4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic Books.

Hohwy, J. (2013). *The predictive mind* (First edition). Oxford University Press.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107–129. https://doi.org/10.1016/0001-6918(67)90011-X

Hunt, E. (2011). *Human intelligence.* (pp. xiv, 507). Cambridge University Press.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Hutchins, E. (2006). *Cognition in the wild* (8. pr). MIT Press.

Hutchinson, J. M. C., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, *69*(2), 97–124. https://doi.org/10.1016/j.beproc.2005.02.019

Icard, T. (n.d.). *Toward Boundedly Rational Analysis*. 6.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jeffrey, R. C. (1996). *The logic of decision* (2. ed). Univ. of Chicago Press.

Just, D. R. (2020). John Kay and Mervyn King: Radical Uncertainty: Decision-Making Beyond the Numbers. *Administrative Science Quarterly*, *65*(3), NP24–NP26. https://doi.org/10.1177/0001839220918981

Kagel, J. H., & Roth, A. E. (2020). *The handbook of experimental economics, volume 2*. Princeton university press.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, *93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin Books.

Kahneman, D., Diener, E., & Schwarz, N. (1999). *Well-Being*. Russell Sage Foundation; JSTOR. http://www.jstor.org/stable/10.7758/9781610443258

Kahneman, D., & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (1st ed., pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, *98*(6), 1325–1348. JSTOR.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*. 15.

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

Kahneman, D., & Tversky, A. (n.d.). *ON THE PSYCHOLOGY OF PREDICTION*. 15.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, *11*(2), 123–141. https://doi.org/10.1016/0010-0277(82)90022-1

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

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

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. https://doi.org/10.3758/BF03196323

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66

Kanfer, R., & Ackerman, P. L. (2004). Aging, Adult Development, and Work Motivation. *The Academy of Management Review*, *29*(3), 440. https://doi.org/10.2307/20159053

Katsikopoulos, K. V. (2014). Bounded rationality: The two cultures. *Journal of Economic Methodology*, *21*(4), 361–374. https://doi.org/10.1080/1350178X.2014.965908

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press.

Kauffman, S., & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, *128*(1), 11–45. https://doi.org/10.1016/S0022-5193(87)80029-2

Kay, J. A., & King, M. A. (2021). *Radical Uncertainty: Decision-Making Beyond the Numbers*.

Kent, J., Leitner, M., & Curtis, A. (2006). Evaluating the usefulness of functional distance measures when calibrating journey-to-crime distance decay functions. *Computers, Environment and Urban Systems*, *30*(2), 181–200. https://doi.org/10.1016/j.compenvurbsys.2004.10.002

Keynes, J. M. (1937). The General Theory of Employment. *The Quarterly Journal of Economics*, *51*(2), 209. https://doi.org/10.2307/1882087

Kivisto, P. (2013). *Illuminating Social Life: Classical and Contemporary Theory Revisited*. SAGE Publications, Inc. https://doi.org/10.4135/9781506335483

Klaes, M., & Sent, E.-M. (2005). A Conceptual History of the Emergence of Bounded Rationality. *History of Political Economy*, *37*(1), 27–59. https://doi.org/10.1215/00182702-37-1-27

Klein, G. (2008). Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Klein, G. (2015). A naturalistic decision making perspective on studying intuitive decision making. *Journal of Applied Research in Memory and Cognition*, *4*(3), 164–168. https://doi.org/10.1016/j.jarmac.2015.07.001

Klein, G. A. (1998). *Sources of power: How people make decisions*. MIT Press.

Klein, G. A. (2009). *Streetlights and shadows: Searching for the keys to adaptive decision making*. MIT Press.

Klein, G. A. (2013). *Seeing what others don't: The remarkable ways we gain insights* (First edition). PublicAffairs.

Kleinmuntz, B., & Cattell, R. B. (1968). *Formal representation of human judgment* (Symposium on Cognition (3rd 1967 Carnegie Institute of Technology Carnegie-Mellon University, Ed.). Wiley.

Kliger, D., van den Assem, M. J., & Zwinkels, R. C. J. (2014). Empirical behavioral finance. *Journal of Economic Behavior & Organization*, *107*, 421–427. https://doi.org/10.1016/j.jebo.2014.10.012

Knauff, M., & Spohn, W. (Eds.). (2021). *The handbook of rationality*. The MIT Press.

Knight, F. H. (1921). *Risk, uncertainty and profit*. University of Chicago Press.

Kobayashi, K., Lee, S., Filipowicz, A. L. S., McGaughey, K. D., Kable, J. W., & Nassar, M. R. (n.d.). *Dynamic Representation of the Subjective Value of Information*. 31.

Kornblith, H., & Foley, R. (1990). The Theory of Epistemic Rationality. *The Philosophical Review*, *99*(1), 131. https://doi.org/10.2307/2185217

Kozyreva, A., & Hertwig, R. (2019). The interpretation of uncertainty in ecological rationality. *Synthese*, *198*(2), 1517–1547. https://doi.org/10.1007/s11229-019-02140-w

Kozyreva, A., Pachur, T., Pleskac, T. J., & Hertwig, R. (2019). *Interpreting uncertainty: A brief history of not knowing*.

Krampen, M., Oehler, K., Posner, R., Sebeok, T. A., & von Uexküll, T. (Eds.). (1987). *Classics of Semiotics*. Springer US. https://doi.org/10.1007/978-1-4757-9700-8

Kreps, D. M. (1988). *Notes on the theory of choice*. Westview Press.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Viking.

Laibson, D., & List, J. A. (2015). Principles of (Behavioral) Economics. *American Economic Review*, *105*(5), 385–390. https://doi.org/10.1257/aer.p20151047

Laudan, L. (1971). William Whewell on the Consilience of Inductions. *The Monist*, *55*(3).

Lee, C. (1949). Essentials of psychological testing. *Science Education*, *35*(4), 227–227. https://doi.org/10.1002/sce.3730350432

Lehman, J., Gordon, J., Jain, S., Ndousse, K., Yeh, C., & Stanley, K. O. (2022). *Evolution through Large Models*. https://doi.org/10.48550/ARXIV.2206.08896

Levine, N., & Block, R. (2011). Bayesian Journey-to-Crime Estimation: An Improvement in Geographic Profiling Methodology. *The Professional Geographer*, *63*(2), 213–229. https://doi.org/10.1080/00330124.2010.547152

Levinthal, D. A. (1997). Adaptation on Rugged Landscapes. *Management Science*, *43*(7), 934–950. https://doi.org/10.1287/mnsc.43.7.934

Lewis, M. (2017). *The undoing project: A friendship that changed our minds* (First edition). W.W. Norton & Company.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational Rationality: Linking Mechanism and Behavior Through Bounded Utility Maximization. *Topics in Cognitive Science*, *6*(2), 279–311. https://doi.org/10.1111/tops.12086

Lieder, F. (n.d.). *When to use which heuristic: A rational solution to the strategy selection problem*. 6.

Lieder, F., & Goodman, N. D. (n.d.). *Reverse-Engineering Resource-Efficient Algorithms*. 8.

Lieder, F., & Griffiths, T. L. (2015). *Model-based strategy selection learning*. https://doi.org/10.13140/RG.2.2.26401.12640

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. https://doi.org/10.1017/S0140525X1900061X

Lieder, F., Griffiths, T. L., Huys, Q. J. M., & Goodman, N. D. (n.d.). *Testing alternative models of anchoring and adjustment*. 12.

Lieder, F., Griffiths, T. L., M. Huys, Q. J., & Goodman, N. D. (2018a). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, *25*(1), 322–349. https://doi.org/10.3758/s13423-017-1286-8

Lieder, F., Griffiths, T. L., M. Huys, Q. J., & Goodman, N. D. (2018b). Empirical evidence for resource-rational anchoring and adjustment. *Psychonomic Bulletin & Review*, *25*(2), 775–784. https://doi.org/10.3758/s13423-017-1288-6

Lieder, F., Krueger, P. M., Callaway, F., & Griffiths, T. (2017). *A reward shaping method for promoting metacognitive learning* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/qj346

Lieder, F., Krueger, P. M., & Griffiths, T. L. (n.d.). *An automatic method for discovering rational heuristics for risky choice*. 6.

Lindley, D. V. (2014). *Understanding uncertainty* (Revised edition). John Wiley & Sons, Inc.

Loewenstein, G., Read, D., & Baumeister, R. F. (Eds.). (2003). *Time and decision: Economic and psychological perspectives on intertemporal choice*. Russell Sage Foundation.

Lorenz, K. (1970). *Studies in Animal and Human Behaviour. Volume I.* https://doi.org/10.4159/harvard.9780674430389

Luce, R. D. (2014). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Psychology Press.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. Dover Publications.

Macchi, L., & Bagassi, M. (2012). Intuitive and analytical processes in insight problem solving: A psycho-rhetorical approach to the study of reasoning. *Mind & Society*, *11*(1), 53–67. https://doi.org/10.1007/s11299-012-0103-3

Macchi, L., & Bagassi, M. (2019). The Argumentative and the Interpretative Functions of Thought. *Teorema: Revista Internacional de Filosofía*, *38*(1), 87–96. JSTOR.

Machina, M. J., & Schmeidler, D. (1992). A More Robust Definition of Subjective Probability. *Econometrica*, *60*(4), 745. https://doi.org/10.2307/2951565

Manktelow, K., & Galbraith, N. (2012). *Thinking and Reasoning* (0 ed.). Psychology Press. https://doi.org/10.4324/9780203115466

Manktelow, K. I., & Chung, M. C. (2004). *Psychology of reasoning: Theoretical and historical perspectives*. Psychology Press. http://www.myilibrary.com?id=14253

Manktelow, K. I., & Over, D. E. (Eds.). (1993). *Rationality: Psychological and philosophical perspectives*. Routledge.

Markowitz, H. (1952). PORTFOLIO SELECTION*. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Maroiu, C., & Maricuţoiu, L. P. (2021). Is the transitivity of choices a proper measure of rationality? *Current Psychology*, *40*(6), 2654–2664. https://doi.org/10.1007/s12144-019-00187-z

Marr, D. (1976). *Artificial Intelligence—A Personal View*. North-Holland Publishing Company.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, *52*(1), 29–71. https://doi.org/10.1023/A:1015516217425

Martignon, L., & Schmitt, M. (1999). *Simplicity and Robustness of Fast and Frugal Heuristics*. 29.

Martignon, L., Vitouch, O., Takezawa, M., & Forster, M. R. (2005). Naive and Yet Enlightened: From Natural Frequencies to Fast and Frugal Decision Trees. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (pp. 189–211). John Wiley & Sons, Ltd. https://doi.org/10.1002/047001332X.ch10

Massimi, M., & McCoy, C. D. (Eds.). (2019). *Understanding perspectivism: Scientific and methodological prospects*. Taylor and Francis.

Mastrogiorgio, A., Felin, T., Kauffman, S., & Mastrogiorgio, M. (2022). More Thumbs Than Rules: Is Rationality an Exaptation? *Frontiers in Psychology*, *13*, 805743. https://doi.org/10.3389/fpsyg.2022.805743

McCarthy, J., & Hayes, P. J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. *Edinburgh University Press*, *Machine Intelligence 4*.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

McShane, M., Bringsjord, S., Hendler, J., Nirenburg, S., & Sun, R. (2019). A Response to Núñez et al.'s (2019) "What Happened to Cognitive Science?" *Topics in Cognitive Science*, *11*(4), 914–917. https://doi.org/10.1111/tops.12458

Mele, A. R., & Rawling, P. (Eds.). (2004). *The Oxford handbook of rationality*. Oxford University Press.

Mellers, B. A., Schwartz, A., & Cooke, A. D. J. (n.d.). *JUDGMENT AND DECISION MAKING*. 32.

Milli, S., Lieder, F., & Griffiths, T. L. (n.d.). *When does Bounded-Optimal Metareasoning Favor Few Cognitive Systems?* 7.

Milli, S., Lieder, F., & Griffiths, T. L. (2021). A rational reinterpretation of dual-process theories. *Cognition*, *217*, 104881. https://doi.org/10.1016/j.cognition.2021.104881

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. PubMed. https://doi.org/10.1177/0963721411429458

Mohnert, F., Pachur, T., & Lieder, F. (2019). *What's in the Adaptive Toolbox and How Do People Choose From It? Rational Models of Strategy Selection in Risky Choice*. https://doi.org/10.13140/RG.2.2.23098.11200

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Mulhall, S., Swift, A., & Rawls, J. (1994). Political Liberalism. *The Philosophical Quarterly*, *44*(177), 542. https://doi.org/10.2307/2220257

Nassar, J., Sokol, P. A., Chung, S., Harris, K. D., & Park, I. M. (2020). On 1/n neural representation and robustness. *ArXiv:2012.04729 [Cs]*. http://arxiv.org/abs/2012.04729

Nassar, M. R., Waltz, J. A., Albrecht, M. A., Gold, J. M., & Frank, M. J. (2021). All or nothing belief updating in patients with schizophrenia reduces precision and flexibility of beliefs. *Brain*, *144*(3), 1013–1029. https://doi.org/10.1093/brain/awaa453

Navon, D. (1989). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system II. application to the field of attention. *European Journal of Cognitive Psychology*, *1*(3), 215–238. https://doi.org/10.1080/09541448908403082

Nickles, T. (2018). Bounded rationality, scissors, crowbars, and pragmatism: Reflections on Herbert Simon. *Mind & Society*, *17*(1–2), 85–96. https://doi.org/10.1007/s11299-019-00206-3

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012a). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012b). “Intelligence: New findings and theoretical developments”: Correction to Nisbett et al. (2012). *American Psychologist*, *67*(2), 129–129. https://doi.org/10.1037/a0027240

Nobandegani, A. S. (2017). *The Minimalist Mind*. McGill University.

Nozick, R. (1993). *The nature of rationality* (Third printing, first Princeton paperback printing). Princeton University Press.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, *3*(8), 782–791. https://doi.org/10.1038/s41562-019-0626-2

Oaksford, M. (1995). Information gain explains relevance which explains the selection task. *Cognition*, *57*(1), 97–108. https://doi.org/10.1016/0010-0277(95)00671-K

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Psychology Press. http://www.myilibrary.com?id=6244

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*(8), 349–357. https://doi.org/10.1016/S1364-6613(00)01699-5

Okasha, S. (2016). ON THE INTERPRETATION OF DECISION THEORY. *Economics and Philosophy*, *32*(3), 409–433. https://doi.org/10.1017/S0266267115000346

Okasha, S., & Binmore, K. (Eds.). (2012). *Evolution and Rationality: Decisions, Co-operation and Strategic Behaviour* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511792601

O’Sullivan, F. (1986). A Statistical Perspective on Ill-Posed Inverse Problems. *Statistical Science*, *1*(4). https://doi.org/10.1214/ss/1177013525

Otworowska, M., Blokpoel, M., Sweers, M., Wareham, T., & van Rooij, I. (2018). Demons of Ecological Rationality. *Cognitive Science*, *42*(3), 1057–1066. https://doi.org/10.1111/cogs.12530

Over, D. (2004). Rationality and the Normative/Descriptive Distinction. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 1–18). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470752937.ch1

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*(6), 972–987. https://doi.org/10.1037/0022-3514.76.6.972

Pashler, H. (1998). *Attention.* (pp. viii, 407). Psychology Press/Erlbaum (UK) Taylor & Francis.

Pawitan, Y., & Lee, Y. (2021). On rereading Savage. *ArXiv:2110.15012 [Math, Stat]*. http://arxiv.org/abs/2110.15012

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*(1), 29–46. https://doi.org/10.1037/h0024722

Petracca, E. (2015). *A tale of paradigm clash: Simon, situated cognition and the interpretation of bounded rationality*. 37.

Petracca, E. (2017). A cognition paradigm clash: Simon, *situated cognition* and the interpretation of bounded rationality. *Journal of Economic Methodology*, *24*(1), 20–40. https://doi.org/10.1080/1350178X.2017.1279742

Petracca, E. (2021). Embodying Bounded Rationality: From Embodied Bounded Rationality to Embodied Rationality. *Frontiers in Psychology*, *12*, 710607. https://doi.org/10.3389/fpsyg.2021.710607

Pinker, S. (2021). *Rationality: What it is, why it seems scarce, why it matters*. Allen Lane, an imprint of Penguin Books.

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, *143*(5), 2000–2019. https://doi.org/10.1037/xge0000013

Raab, M. (2004). Intelligence as Smart Heuristics. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and Intelligence* (1st ed., pp. 188–207). Cambridge University Press. https://doi.org/10.1017/CBO9780511607073.011

Raab, M., & Gigerenzer, G. (2015). The power of simplicity: A fast-and-frugal heuristics approach to performance science. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01672

Rakow, T. (2010). *Risk, uncertainty and prophet: The psychological insights of Frank H. Knight*. 17.

Ramsey, F. P. (1926). *TRUTH AND PROBABILITY*. 41.

Read, D., & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, *1*(1), 34–49. https://doi.org/10.1037/1076-898X.1.1.34

Reed, S. K. (2016). The Structure of Ill-Structured (and Well-Structured) Problems Revisited. *Educational Psychology Review*, *28*(4), 691–716. https://doi.org/10.1007/s10648-015-9343-1

Rich, P. (2016). Axiomatic and ecological rationality: Choosing costs and benefits. *Erasmus Journal for Philosophy and Economics*, *9*(2), 90. https://doi.org/10.23941/ejpe.v9i2.231

Rich, P., Blokpoel, M., de Haan, R., Otworowska, M., Sweers, M., Wareham, T., & van Rooij, I. (2021). Naturalism, tractability and the adaptive toolbox. *Synthese*, *198*(6), 5749–5784. https://doi.org/10.1007/s11229-019-02431-2

Roli, A., Jaeger, J., & Kauffman, S. (2021). *How organisms come to know the world: Fundamental limits on artificial general intelligence* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/yfmt3

Rolla, G. (2021). Reconceiving rationality: Situating rationality into radically enactive cognition. *Synthese*, *198*(S1), 571–590. https://doi.org/10.1007/s11229-019-02362-y

Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., Davison, C., Demić, E., Esteban-Serna, C., Friedemann, M., Gibson, S. P., Jarke, H., Karakasheva, R., Khorrami, P. R., Kveder, J., Andersen, T. L., Lofthus, I. S., McGill, L., Nieto, A. E., … Folke, T. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, *4*(6), 622–633. https://doi.org/10.1038/s41562-020-0886-x

Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The cost structure of sensemaking. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*, 269–276. https://doi.org/10.1145/169059.169209

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (Fourth edition, global edition). Pearson.

Russell, S. J., & Subramanian, D. (1995). Provably Bounded-Optimal Agents. *Journal of Artificial Intelligence Research*, *2*, 575–609. https://doi.org/10.1613/jair.133

Russell, S. J., & Wefald, E. (1991). *Do the right thing: Studies in limited rationality*. MIT Press.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive Functioning as a Potential Mediator of Age-Related Cognitive Decline in Normal Adults. *Journal of Experimental Psychology: General*, *132*(4), 566–594. https://doi.org/10.1037/0096-3445.132.4.566

Samuels, R., Stich, S., & Bishop, M. (2002). Ending the Rationality Wars How to Make Disputes about Human Rationality Disappear. In R. Elio (Ed.), *Common Sense, Reasoning, and Rationality* (pp. 236–268). Oxford University Press. https://doi.org/10.1093/0195147669.003.0011

Savage, L. J. (1972). *The foundations of statistics* (2d rev. ed). Dover Publications.

Schelling, T. C. (1984). Self-Command in Practice, in Policy, and in a Theory of Rational Choice. *The American Economic Review*, *74*(2), 1–11. JSTOR.

Schelling, T. C. (1993). *Choice and consequence: Perspectives of an errant economist*. Harvard Univ. Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schooler, J. W. (2014). Turning the Lens of Science on Itself: Verbal Overshadowing, Replication, and Metascience. *Perspectives on Psychological Science*, *9*(5), 579–584. https://doi.org/10.1177/1745691614547878

Schurz, G., & Hertwig, R. (2019). Cognitive Success: A Consequentialist Account of Rationality in Cognition. *Topics in Cognitive Science*, *11*(1), 7–36. https://doi.org/10.1111/tops.12410

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*(2), 309–316. https://doi.org/10.1037/0033-2909.105.2.309

Selten, R. (1990). Bounded Rationality. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift Für Die Gesamte Staatswissenschaft*, *146*(4), 649–658. JSTOR.

Sent, E.-M. (2002). Bounded Rationality: The Adaptive Toolbox. *Journal of Economic Issues*, *36*(1), 232–234. https://doi.org/10.1080/00213624.2002.11506464

Sent, E.-M. (2004). The legacy of Herbert Simon in game theory. *Journal of Economic Behavior & Organization*, *53*(3), 303–317. https://doi.org/10.1016/j.jebo.2002.10.006

Sent, E.-M. (2005). Simplifying Herbert Simon. *History of Political Economy*, *37*(2), 227–232. https://doi.org/10.1215/00182702-37-2-227

Sent, E.-M. (2018). Rationality and bounded rationality: You can't have one without the other. *The European Journal of the History of Economic Thought*, *25*(6), 1370–1386. https://doi.org/10.1080/09672567.2018.1523206

Sharp, P. B., Fradkin, I., & Eldar, E. (2022). Hierarchical inference as a source of human biases. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-022-01020-0

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, *69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Silver, D. (2015). *Reinforcement Learning Course by David Silver—Lecture 1: Introduction to Reinforcement Learning (Deepmind) [Video]*. https://youtu.be/2pWv7GOvuf0

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2017). Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. *ArXiv:1712.01815 [Cs]*. http://arxiv.org/abs/1712.01815

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, 103535. https://doi.org/10.1016/j.artint.2021.103535

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99. https://doi.org/10.2307/1884852

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Simon, H. A. (1987). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*. Garland Pub.

Simon, H. A. (1997). *Administrative behavior: A study of decision-making processes in administrative organizations* (4th ed). Free Press.

Simon, H. A., & Newell, A. (1958). *Report on a General Problem-Solving Program*.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, *26*(2), 145–159. https://doi.org/10.1037/h0030806

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1

Skyrms, B. (2000). *Choice and chance: An introduction to inductive logic* (4th ed). Wadsworth/Thomson Learning.

Sloman, S. (1996). The Empirical Case For Two Systems of Reasoning. *Psychological Bulletin*, *119*, 3. https://doi.org/10.1037/0033-2909.119.1.3

Sloman, S. A. (2002). Two Systems of Reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (1st ed., pp. 379–396). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.024

Slovic, P. (2000). *The perception of risk*. Earthscan Publications.

Slovic, P., & Finucane, M. (2002). *The affect heuristic. Heuristics and Biases: The Psychology of Intuitive Judgment. DG Gilovich and D. Kahneman*.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis*, *24*(2), 311–322. https://doi.org/10.1111/j.0272-4332.2004.00433.x

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1984). Behavioral decision theory perspectives on risk and safety. *Acta Psychologica*, *56*(1), 183–203. https://doi.org/10.1016/0001-6918(84)90018-0

Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science*, *19*(6), 368–373. https://doi.org/10.1002/bs.3830190603

Slovic Paul. (1987). Perception of Risk. *Science*, *236*(4799), 280–285. https://doi.org/10.1126/science.3563507

Smith, V. L. (2003). Constructivist and Ecological Rationality in Economics. *American Economic Review*, *93*(3), 465–508. https://doi.org/10.1257/000282803322156954

Smith, V. L. (2008). *Rationality in economics: Constructivist and ecological forms*. Cambridge University Press.

Sperber, D., & Wilson, D. (2001). *Relevance: Communication and cognition* (2nd ed). Blackwell Publishers.

Spiliopoulos, L., & Hertwig, R. (2020). A map of ecologically rational heuristics for uncertain strategic worlds. *Psychological Review*, *127*(2), 245–280. https://doi.org/10.1037/rev0000171

Sprenger, J., & Hartmann, S. (2019). *Bayesian philosophy of science: Variations on a theme by the Reverend Thomas Bayes* (First edition). Oxford University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford University Press.

Stanovich, K. E. (2012). *On the Distinction Between Rationality and Intelligence: Implications for Understanding Individual Differences in Reasoning*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0022

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Lawrence Erlbaum Associates.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale Univ. Press.

Stanovich, K. E. (2010). *Decision making and rationality in the modern world*. Oxford University Press.

Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning*, *19*(1), 1–26. https://doi.org/10.1080/13546783.2012.713178

Stanovich, K. E. (2016). The Comprehensive Assessment of Rational Thinking. *Educational Psychologist*, *51*(1), 23–34. https://doi.org/10.1080/00461520.2015.1125787

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Stanovich, K. E., Toplak, M. E., & West, R. F. (2016). *The rationality quotient: Toward a test of rational thinking*. MIT Press.

Stanovich, K. E., & West, R. F. (1999). Discrepancies Between Normative and Descriptive Models of Decision Making and the Understanding/Acceptance Principle. *Cognitive Psychology*, *38*(3), 349–385. https://doi.org/10.1006/cogp.1998.0700

Stanovich, K. E., & West, R. F. (2000a). Advancing the rationality debate. *Behavioral and Brain Sciences*, *23*(5), 701–717. https://doi.org/10.1017/S0140525X00623439

Stanovich, K. E., & West, R. F. (2000b). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435

Stanovich, K. E., & West, R. F. (2002). Individual Differences in Reasoning: Implications for the Rationality Debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (1st ed., pp. 421–440). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.026

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*(3), 225–247. https://doi.org/10.1080/13546780600780796

Stearns, S. C. (2000). Daniel Bernoulli (1738): Evolution and economics under risk. *Journal of Biosciences*, *25*(3), 221–228. https://doi.org/10.1007/BF02703928

Stein, E. (1997). *Without Good Reason*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237730.001.0001

Sternberg, R. J. (Ed.). (2018). *The Nature of Human Intelligence* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316817049

Sternberg, R. J., Grigorenko, E. L., & Zhang, L. (2008). Styles of Learning and Thinking Matter in Instruction and Assessment. *Perspectives on Psychological Science*, *3*(6), 486–506. https://doi.org/10.1111/j.1745-6924.2008.00095.x

Sternberg, R., & Sternberg, K. (2016). *Cognitive psychology* (7th edition). Cengage Learning.

Sturm, T. (2012). The "Rationality Wars" in Psychology: Where They Are and Where They Could Go. *Inquiry*, *55*(1), 66–81. https://doi.org/10.1080/0020174X.2012.643628

Sturm, T. (2019). Formal versus Bounded Norms in the Psychology of Rationality: Toward a Multilevel Analysis of Their Relationship. *Philosophy of the Social Sciences*, *49*(3), 190–209. https://doi.org/10.1177/0048393119842786

Sunstein, C. R. (1999). Behavioral law and economics: A progress report. *Am. L. & Econ. Rev.*, *1*, 115.

Sunstein, C. R. (2018). *The cost-benefit revolution*. The MIT Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. 352.

Taleb, N. N. (2011). A Map and Simple Heuristic to Detect Fragility, Antifragility, and Model Error. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1864633

Taleb, N. N. (2018). *Skin in the game: Hidden asymmetries in daily life* (First edition). Random House.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Tetlock, P. E., & Mellers, B. A. (2002). The Great Rationality Debate. *Psychological Science*, *13*(1), 94–99. https://doi.org/10.1111/1467-9280.00418

Thagard, P., & Findlay, S. (2012). *The cognitive science of science: Explanation, discovery, and conceptual change*. MIT Press.

Thaler, R. (n.d.). *TOWARD A POSITIVE THEORY OF CONSUMER CHOICE*. 22.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* (pp. x, 293). Yale University Press.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Belknap Press of Harvard University Press.

Tinbergen, N. (1953). The evolution of mating behavior patterns. *Evolution*, *7*(4), 391–391. https://doi.org/10.1111/j.1558-5646.1953.tb00102.x

Titelbaum, M. G. (2013). *Quitting certainties: A Bayesian framework modeling degrees of belief* (1st ed). Oxford University Press.

Titelbaum, M. G. (2022). *Fundamentals of Bayesian epistemology 1: Introducing credences*. Oxford University Press.

Titelbaum, M. G. (2022). *Fundamentals of Bayesian Epistemology 2: Arguments, challenges, alternatives.* Oxford University Press.

Todd, M., Fiddick, L., & Krauss, S. (2000). Ecological rationality and its contents. *Thinking & Reasoning*, *6*(4), 375–384. https://doi.org/10.1080/135467800750038184

Todd, P. M., & Brighton, H. (2016). Building the Theory of Ecological Rationality. *Minds and Machines*, *26*(1–2), 9–30. https://doi.org/10.1007/s11023-015-9371-0

Todd, P. M., & Gigerenzer, G. (2007). Environments That Make Us Smart: Ecological Rationality. *Current Directions in Psychological Science*, *16*(3), 167–171. https://doi.org/10.1111/j.1467-8721.2007.00497.x

Todd, P. M., & Gigerenzer, G. (Eds.). (2012). *Ecological rationality: Intelligence in the world*. Oxford University Press.

Todd, P. M., Hills, T. T., & Robbins, T. W. (Eds.). (2012). *Cognitive search: Evolution, algorithms, and the brain*. MIT Press.

Tomlin, D., Rand, D. G., Ludvig, E. A., & Cohen, J. D. (2015). The evolution and devolution of cognitive control: The costs of deliberation in a competitive world. *Scientific Reports*, *5*(1), 11002. https://doi.org/10.1038/srep11002

Tooby, J., & Cosmides, L. (2005). Evolutionary psychology: Conceptual foundations. *The Handbook of Evolutionary Psychology*.

Toplak, M. E., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making*, *20*(2), 103–124. https://doi.org/10.1002/bdm.544

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). *Judgment under Uncertainty: Heuristics and Biases*. *185*, 10.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1986). The Framing of Decisions and the Evaluation of Prospects. In *Studies in Logic and the Foundations of Mathematics* (Vol. 114, pp. 503–520). Elsevier. https://doi.org/10.1016/S0049-237X(09)70710-4

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

TVERSKY, A., & KAHNEMAN, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. JSTOR.

van der Putten, P., & van Someren, M. (2004). A Bias-Variance Analysis of a Real World Learning Problem: The CoIL Challenge 2000. *Machine Learning*, *57*(1/2), 177–195. https://doi.org/10.1023/B:MACH.0000035476.95130.99

van Rooij, I., Wright, C. D., Kwisthout, J., & Wareham, T. (2018). Rational analysis, intractability, and the prospects of 'as if'-explanations. *Synthese*, *195*(2), 491–510. https://doi.org/10.1007/s11229-014-0532-0

Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

Vervaeke, J., & Ferraro, L. (2013). Relevance, Meaning and the Cognitive Science of Wisdom. In M. Ferrari & N. M. Weststrate (Eds.), *The Scientific Study of Personal Wisdom* (pp. 21–51). Springer Netherlands. https://doi.org/10.1007/978-94-007-7987-7_2

Vervaeke, J., Lillicrap, T. P., & Richards, B. A. (2012). Relevance Realization and the Emerging Framework in Cognitive Science. *Journal of Logic and Computation*, *22*(1), 79–99. https://doi.org/10.1093/logcom/exp067

Viale, R. (Ed.). (2021). *Routledge handbook of bounded rationality*. Routledge.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (60th anniversary ed). Princeton University Press.

von Uexküll, J. (1921). *Umwelt und Innenwelt der Tiere*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-24819-5

Von Uexküll, J. (1992). A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, *89*(4). https://doi.org/10.1515/semi.1992.89.4.319

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science*, *38*(4), 599–637. https://doi.org/10.1111/cogs.12101

Wallin, A. (2013). A peace treaty for the rationality wars? External validity and its relation to normative and descriptive theories of rationality. *Theory & Psychology*, *23*(4), 458–478. https://doi.org/10.1177/0959354313489369

Watson, W. (1990). STRATEGY, RATIONALITY AND INFERENCE: THE POSSIBILITY OF SYMBOLIC PERFORMANCES. *Sociology*, *24*(3), 485–498. JSTOR.

Weber, M. (2013). *Economy and society. volume 2: Economy and society: an outline of interpretive sociology / Max Weber ; edited by Guenther Roth and Claus Wittich ; with a new foreword by Guenther Roth* (G. Roth & C. Wittich, Eds.). University of California Press.

Weeks, D. (1990). Paradoxes of gambling behaviour. *Journal of Psychosomatic Research*, *34*(4), 477–478. https://doi.org/10.1016/0022-3999(90)90076-G

Weinan, E. (2021). The Dawning of a New Era in Applied Mathematics. *Notices of the American Mathematical Society*, *68*(04), 1. https://doi.org/10.1090/noti2259

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

Wiener, N. (2007). *Cybernetics or control and communication in the animal and the machine* (2. ed., reprint). MIT Press.

Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences—Cognition*. Cambridge University Press.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Harvard University Press.

Wolfram, S. (2002). *A new kind of science*. Wolfram Media.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1109/4235.585893

Wolpert, D., & Macready, W. G. (1995). *No Free Lunch Theorems for Search*.

Woodford, M. (2014). Stochastic Choice: An Optimizing Neuroeconomic Model. *American Economic Review*, *104*(5), 495–500. https://doi.org/10.1257/aer.104.5.495

Wright, G., & Ayton, P. (Eds.). (1994). *Subjective probability*. Wiley.

Zajonc, R. B. (1998). Emotions. In *The handbook of social psychology, Vols. 1-2, 4th ed.* (pp. 591–632). McGraw-Hill.