This is a section of doi:10.7551/mitpress/14668.001.0001

Insolvent

How to Reorient Computing for Just Sustainability

By: Christoph Becker

Citation:

Insolvent: How to Reorient Computing for Just Sustainability
By: Christoph Becker
DOI: 10.7551/mitpress/14668.001.0001
ISBN (electronic): 9780262374668
Publisher: The MIT Press
Published: 2023

The MIT Press

7

PEOPLE ARE MORE THAN RATIONAL

BEWARE THE NORMATIVE FALLACY

It's strange how often the critics of artificial intelligence object to the wrong thing . . . they are horrified at the suggestion that computers can think, whereas they should be horrified at the suggestion that people are information processors.
—Churchman (1979a, 124)

The design and development of a computational system is full of tradeoffs. Software engineers, architects, programmers, testers, user interface designers, project managers, and many others must work in concert with all sorts of stakeholders to navigate design options that shape the system they are making. Because of this, disciplines such as software engineering treat decision-making as a central part of their methodological focus. For example, software architecture is a decision-centric discipline (van Vliet and Tang 2016). Eoin Woods, a leading practitioner and writer, describes it as "the set of decisions which, if made incorrectly, will cause your project to be cancelled" (Bass 2013, 25). Decision-making is similarly central to requirements engineering (Aurum and Wohlin 2003).

How do people make all these decisions? Without understanding what happens in engineering and design practice, we can hardly hope to improve it.[1] Put simply, engineering and design disciplines *prescribe* what people should do and why—they develop normative frameworks such as methods that define what should be done and how. In contrast, studies of behavior *describe* and explain what people do in practice. In behavioral studies *of* engineering and design, these two modes overlap, because they strive to describe and explain what happens in practice in order to prescribe what people could do better, and how, for some standard of evaluation. For example, many researchers in software engineering design new artifacts such as methods and tools and deploy them into industrial contexts, then study how they impact performance. What type of knowledge should provide the foundation of such studies?

Behavioral researchers collect data about professional practice. When they organize this empirical part of their study, they often rely on the toolbox of theories they used to design the methods. But the theories that *describe* what people do are different from the engineering methods that *prescribe* what they should do. When it comes to how people make decisions, these theories in fact carry mutually incompatible assumptions.[2] As a result, the tension between description and prescription can lead behavioral researchers to misunderstand practice in subtle but important ways. When researchers misappropriate normative and prescriptive theories that lack descriptive validity for descriptive purposes, they commit a *normative fallacy*. Their findings may appear persuasive, but they are invalid and they will mislead us.

This chapter retraces the tradition of decision-making research reflected in the myth of rational decisions to outline what it misses and show how it has misled the computing field. The historical view will help illustrate how the myth of rational decision-making lives on in systems design research and practice. I focus on empirical behavioral research of engineering practice to make my case. The argument draws parallels and lessons from developments in other disciplines shaped, just like computing, by the domineering influence of the rationalistic tradition. By exploring how these fields extricated themselves from a singular focus on rationalistic theories, we will see why computing needs to follow suit and how to reorient it.

RATIONAL CHOICES, REASONABLE DECISIONS, WISE JUDGMENTS

The rationalistic tradition dominates the attention to decision-making in computing. When and if disciplines such as software engineering define decision-making, they do so within this tradition, as in this example mentioned in chapter 3: "to make a decision, a situation is assessed against a set of characteristics or attributes, also called criteria" (Filho, Pinheiro, and Albuquerque 2016).

This definition of decision-making as a selection from a predefined enumerated set of options is common outside of computing too: The APA defines decision-making as "the cognitive process of choosing between two or more alternatives" (APA 2020a), which makes it indistinguishable from choice: "an act of selecting or making a decision when faced with two or more possibilities" (*Oxford English Dictionary* 2020a). But as we will see, people often arrive at a decision without performing a choice. The interdisciplinary area of judgment and decision-making (JDM) explores how humans make complex decisions and judgments (Keren and Wu 2015). Its perspectives range from psychology and social psychology to behavioral economics, sociology, and neuroscience. While the terms choice and decision are sometimes used interchangeably, their meaning differs significantly:

1. A *decision* arises in a situation in which someone could conceivably make different commitments on how to proceed. It is a "conclusion or resolution reached after consideration" (*Oxford English Dictionary* 2020b). In naturalistic decision-making, making a decision is in fact defined as "committing oneself to a certain course of action" (Lipshitz et al. 2001).
2. A *choice* is a specific type of decision where enumerated options exist from which a selection has to be made.
3. *Judgment*, on the other hand, always carries a broader awareness and attention to sense-making, evaluation, and the formation of a subject's position toward an object of careful consideration, as in "the process of forming an opinion or evaluation by discerning and comparing" (Merriam-Webster 2020b), the "ability to make considered decisions or come to sensible conclusions" (*Oxford English Dictionary* 2020c) or "the capacity to recognize relationships, draw conclusions from evidence, and make critical evaluations of events and people" (APA 2020b). When you face a choice between two options but reflect on the boundaries of the presented decision, reject the framing, and instead pursue a third option, you exercise judgement. You make a different decision (commitment) as a result of your judgment. That is a complex human capacity that invokes reflection and situational awareness, qualities that are markedly absent in the machinery some describe today as "artificial intelligence" (cf. Crawford 2021; B. Smith 2019).

Imagine you are responsible for a software project. You are behind schedule by two weeks. A colleague suggests that instead of developing one of the main functional components, as initially planned, your team could spend this time integrating an open-source library. It would require some customization and there would still be some need for coding, but your colleague thinks it could save up to a month of effort. There is a chance of failure too, of course. What do you do?

In the rationalist tradition, you specify clear and unambiguous assumptions such as the time horizon and budget, enlist decision criteria—either in a list or in more elaborate structures such as utility trees—and then evaluate each alternative on each criterion to compare them and select the best. Cost-benefit analysis is a form of such multicriteria decision analysis (MCDA), architectural tradeoff decision support methods are another. MCDA requires clear assumptions and, at each point in time, treats two of the corners of the facts/value/scope triangle as fixed. It handles uncertainty via probabilities and the incommensurability of various scales by computing a unified value function, often expressed as utility.

Note that the situation has *become a decision* point because your colleague identified that your team could conceivably commit to a different action than initially planned. Otherwise, you may have simply proceeded. By enumerating two options, it has narrowed into a *choice*. But in the discussion that follows, your team may elect to redefine the framing of the decision. They may decide to widen the scope and inquire into other libraries, they may recognize that this library can also address another aspect of the system's functions, and they may choose to challenge the initial framing. For example, one could argue that the time horizon of this project is not a good scope for the decision since the effects, both positive and negative, play out over the entire lifetime of the system. Others may argue that integrating an open-source library has learning value and is fun. And yet others may argue that engaging and perhaps contributing back to the open-source community has altruistic and strategic value too. In other words, they will exercise *judgment*. In doing so, they will iteratively reflect on and discursively reposition claims about facts, values, and scope (see figure 6.2). The team's reasoning will often not fit easily into rigid MCDA frameworks. This is simply because the human capacity for judgment transcends the operations supported by MCDA. If we define decision-making via *only* the operations supplied by MCDA, we lose sight of the human ability to reflect, to judge the situation, and to generate different ways to act.

DECISION-MAKING IN THE RATIONALIST TRADITION

Winograd and Flores already argued that computing needs to "replace the rationalist[] orientation if we want to understand human thought, language, and action" (1986, 26).[3] Their focus was not on specific decisions *in* systems design, but on a general design orientation and paradigm. Despite their argument's influence in human–computer interaction (HCI), the rationalist tradition persists when it comes to questions of decision-making. Why is that the case? Let us take a closer look at the theory's appeal and origin.

According to this tradition, "intelligence is the work of symbol systems," and the brain is simply another example case of an intelligent symbol system (Simon 1996, 23). Judgment does not sit easily within this idea, but decision-making does. Keenly aware of the tension between normative and empirical accounts of rationality, Simon recognized that the rational choice model of *homo economicus*, which assumed perfect information about choices, criteria, and the environment, needed "drastic revision" to make it compatible with "the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist" (Simon 1955, 99). The main change Simon proposed was "taking into account the *simplifications* the choosing organism may deliberately introduce into its model of the situation in order to bring the model within the range of its *computing* capacity" (emphasis added). In other words, the focus is on human thought as a *subset* of computational processing, limited by assumed constraints. The underlying metaphor of the mind as computer is never questioned. Instead, based on admittedly "casual empiricism" (Simon 1955, 104), the rational model is modified to fit the idea of constrained computing capacity:

> Because of the psychological limits of the organism (particularly with respect to computational and predictive ability), actual human rationality-striving can at best be an extremely crude and *simplified approximation* to the kind of global rationality that is implied, for example, by game-theoretical models. (Simon 1955, 101, emphasis added)

Simon's behavioral model of rationality evolves around a set of behavioral alternatives, some of which are considered by the organism, a set of possible outcomes, a subjective value function for each outcome, a set of consequences per alternative, and a probabilistic model of outcomes per alternative. He and his collaborators used this starting point to examine human problem-solving, organizational behavior, and administrative decision-making. The famous studies that laid empirical claims to Simon's theory of bounded rationality examined human problem-solving for highly constrained well-defined problems such as chess puzzles (Newell and Simon 1972). In contrast to the normative modeling exercises of game theory, these studies were descriptive and explanatory: they involved extensive think-aloud protocol analysis of human subjects. But when we read these protocols and analyses today with the benefit of hindsight, it is striking to see how strongly data collection and analysis itself were predicated on the preformed idea of the human mind as a computer. This reified metaphor made it appear natural that human problem-solving is a search in a defined problem space, performed in the mind by an algorithm that processes information obtained from perception and represented in the brain.

SIMON SAYS, OR: HOW REASON LOST ITS MIND

As a leading behavioral economist wrote, "50 years of dominance of the rational choice paradigm . . . has left most important questions unanswered" (Loewenstein, Rick, and Cohen 2008). Why? What happened during those fifty years? This section explores the legacy and deep influence of the rationalist paradigm, again centering on the work of Herbert Simon and his influences. This will allow us to more deeply understand why we should think of rational decision-making as a myth. I often focus in this book on Simon not because he single-handedly created the myth of rational decision-making (he didn't), but because his work is central to it and unique in its wide-ranging influence on the cognitive sciences, artificial intelligence, computer science, psychology, behavioral economics, design, and political science.[4] Each of these disciplines experienced for decades the torque of the computational model of mind, the framework of rational problem-solving as a spatial search in a bounded space, and the axioms of rationalist decision theory. In each discipline, careful work on the margins proved core assumptions wrong, but struggled to be recognized until the evidence became undeniable and the discomfort with existing paradigms too strong to ignore.

In the cognitive sciences, researchers recognized that cognition cannot be explained merely as information processing within the brain, nor is it reducible to computation—intelligence is *not* symbol processing (Maturana and Varela 1992; Varela, Thompson, and Rosch 1991). Objections came from multiple sides, reflecting the fragmented relationship between the disciplines involved in cognitive sciences, including psychology, biology, neuroscience, and sociology. From a sociological perspective, a well-known ethnographic study by Hutchins (1995), followed by others, argued for a shift in focus from the computational mind to a sociotechnical system involving external configurations such as the controls and displays in a cockpit interacting with the pilots. From a biological perspective, experiments in cognition demonstrated that the theory of cognition as representing external reality in the mind—as in a Von Neumann computer, and as a central tenet of the theory of intelligence as symbol manipulation—is empirically incorrect. The brain does not appear to store and manipulate representations of the outside world in neuronal registers through computational processes. The view of *enactive* cognition emphasizes instead how the structural evolution of the embodied mind and its historical coupling with an environment predispose it to act and interact with this environment, continually bringing forth what we experience as the present moment via a process of *autopoiesis*.[5]

In artificial intelligence (AI) research in computer science, the struggle is ongoing, despite early dissent (Weizenbaum 1976; Winograd and Flores 1986; Dreyfus 1972) and Lucy Suchman's influential refutation of the idea that *plans*, a central concept in AI work at the time, work like programs to be run by individual agents. In her detailed studies, it became clear that plans instead are weak resources used by reasonable humans to act meaningfully in their concrete situations. Her shift from plans to "situated action" resonates with Hutchins's studies (Suchman 2006). As late as 2019, Brian Cantwell Smith (2019) still had to explain that intelligence does not merely involve a procedural logic akin to a computer that operates within the given constraints, but also incorporates *judgment*, including the ability to reflect on the given framing as well as the agent's thought process and to make a conscious choice to transcend it, echoing much earlier calls (Weizenbaum 1976). This ability for judgment relies on an accountability to the real world that sets apart lived intelligence from computational machinery. Smith's reference to the continuous interaction of living beings with their environment mirrors the autopoietic theory arising from the biology of cognition and enactive cognitive science. But even today, the appeal of narrow, computationally understood AI remains strong. What so-called AI can do is not in a meaningful sense intelligent. Showcase examples for AI take place in well-defined domains to which computing is well-suited. These are impressive advances but not toward human reasoning. Tragicomical failures by self-driving cars, text generators, and image classifiers should remind us that these machines lack the human ability for reflective judgment.

In decision-making research, spanning behavioral economics, psychology, and later neuroscience, the rationalist research program was long dominant too, despite active discussions of descriptive, prescriptive, and normative perspectives (Bell, Raiffa, and Tversky 1989). While classical economics assumes that humans are rational, behavioral economics essentially assumed the opposite: they are irrational (Harper, Randall, and Sharrock 2016). From a popular perspective, behavioral economists like Kahneman are often portrayed as revolutionaries, but their work remained tethered to normative theories. "By emphasizing the divergence between actual human reasoning and standards of formal rationality such as logic and Bayesian statistics, they implicitly reinforced the normative authority of the latter" (Erickson et al. 2013, 24). But over time, the emphasis similarly shifted from an isolated individual with flawed computing powers to an appreciation of the macro-cognitive system that extends beyond the mind (G. Klein 1998; Hutchins 1995; Thaler, Sunstein, and Balz 2010). This involved important shifts in method too, which we discuss later.

Design-oriented disciplines such as design studies and later design research in HCI and information systems were also heavily influenced by Simon's work on the "sciences of the artificial," which he positioned as "sciences of design," again built on theories of rational problem-solving through satisficing search in defined problem spaces (Simon 1996; Rosner 2018; Dorst 2006). Here too, the burden of proof came to rest with those recognizing the limitations of such an appealing but narrow view. Here, too, the evidence was often of a qualitative, situated nature that was initially sidelined and ignored.[6] As a result, "much contemporary design research, in its pursuit of academic respectability, remains aligned to Simon's broader project, particularly in its definition of design as 'scientific' problem solving. However, the repression of judgment, intuition, experience, and social interaction in Simon's 'logic of design' has had, and continues to have, profound implications for design research and practice" (Huppatz 2015). There are many facets to this dominance, some of which are addressed in previous chapters (see also Rosner 2018). For example, detailed studies of design activity showed that Simon's core assumption of the problem space as existing prior to its exploration is misleading. Instead, designers co-construct and adapt their understanding of problem space and solution space simultaneously—the spaces co-evolve in mutual interaction (Dorst 1995; Dorst and Cross 2001). More generally, design as value-neutral scientific problem solving engenders a paternalistic attitude towards designers as well, as explored in chapter 4. It excludes both the notion of judgment as well as the idea of reflective practice, which evidently happens in design (Schön 1983). The poverty of this concept did not remain unnoticed even then. But Simon's reaction to such counterpoints as wicked problems in design (Buchanan 1992; Cross 1984) was to ignore the terminology and instead propose that "ill structured" problems were not fundamentally distinct from others (Simon 1973). Huppatz (2015) characterizes what I earlier called the evacuation of politics from design by emphasizing the disembodied nature of Simon's designer: "Freed of situated bodies, Simon's 'science of design' failed to engage with designing as a fundamentally social, political, cultural, and embodied activity." In the broader social context, this approach had lasting appeal: "The logic of optimization promises greater predictability and profit while rigorously stripping judgment, intuition, and experience from systems and service design" (Huppatz 2015).

Finally, in Simon's initial home discipline, political science, his work exerted significant influence too and helped shape what later historians called "Cold War Rationality." In *How Reason Almost Lost Its Mind*, Erickson et al. trace how the dominant vision about reason narrowed into a formalized, technical form of reasoning grounded in the formal abstractions of game theory and operations research, a view that explicitly banishes judgment. They describe the resulting view of rationality as a "distinctive combination of stripped-down formalism, economic calculation, optimization, analogical reasoning from experimental microcosms, and towering ambitions" (4) and write:

> What was distinctive about Cold War rationality was the expansion of the domain of rationality at the expense of that of reason, asserting its claims in the loftiest realms of political decision making and scientific method—and sometimes not only in competition with but in downright opposition to reason, reasonableness, and common sense . . . what was Cold War rationality? . . . First of all, this rationality should be formal, and therefore largely independent of personality or context. It frequently took the form of algorithms . . . supposed to provide optimal solutions to given problems, or delineate the most efficient means toward certain given goals. (Erickson et al. 2013, 2–3)

In hindsight, the political forces that shaped this particular narrowing of view within the Cold War context have become much easier to discern, and the streams of ideas much easier to distinguish. Formal, mathematically grounded models of deductive reasoning became not only the normative ideal of how decisions should be made but were also used as the basis for models of how thinking actually works, where they underpinned a long stream of experimental and quasi-experimental research. This research was designed to understand *human* reasoning but anchored on ideas of formal and procedural rationality. Only late in this process did the objections to these flawed assumptions become so strong that they could no longer be ignored.

These accounts of rationality reveal important cross-disciplinary dynamics. There were significant objections within cognitive sciences, psychology, and behavioral economics to the dominant normative narrative and to the false dichotomy that anything not procedurally rational in the narrow sense was supposedly "irrational." But this dissent was hardly recognized and taken up in other fields such as computing or political science. Instead, "if we turn to the applications of the psychology of rationality to policy analysis, hints of dissent within the ranks are very rare. The monolithic definitions of rationality and irrationality inherent to the heuristics-and-biases school of thought still hold sway" (Erickson et al. 2013, 178). Evidently, and unsurprisingly, the same is true for the applications of the psychology of rationality in computing. What exactly is wrong with that?

WHY RATIONALIST DECISION-MAKING IS INADEQUATE

Rationalist models are appealing to researchers because their mathematical formulas promise a rigorous model of human behavior that supports the collection of data, the detection of deviations, and the design of interventions. But it is important to understand that the axioms of rationalist decision theory were not empirical facts, but normative directives. None of them had been established by observation. Empirically, the rationalist theories of decision-making built on these axioms turn out to be invalid in three important ways: assumptions, predictions, and methods.[7]

INVALID ASSUMPTIONS

First, the rationalist theory assumes that preferences are fixed priors, that options are independent from each other, that preference relations are transitive, and that evaluation is independent of irrelevant alternatives. These assumptions are empirically invalid: they do not describe how humans reason (Beach and Lipshitz 1993; Tversky and Kahneman 1986). The assumptions cannot be easily adjusted to fix the theories because they are their foundational axioms without which the theories do not work (Beach and Lipshitz 1993; Shafer 1986). Even more, the underlying assumption that the computer is a reasonable metaphor for the mind has never been verified either—on the contrary, research in the biology of cognition suggests that "the popular metaphor of calling the brain an 'information processing device' is not only ambiguous but patently wrong" (Maturana and Varela 1992, 169). That the brain can *perform* information processing does not reduce it to an information processor.

INCONSISTENT PREDICTIONS

Second, the rationalist predictions of the choices people will make are inconsistent with human behavior. Empirical inconsistencies with the original rational choice theory were known since the 1950s (Beach and Lipshitz 1993; Shafer 1986). To economists, they were long not important because their main focus is not the individual apparatus of cognition and decision-making but the aggregate behavior of the economic system, so it is to some degree defensible for them to abstract away from human reasoning. But for psychologists, the main focus is human reasoning, so the emergence of cognitive psychology and the import of rational theory into the field of psychology meant that the inconsistencies had to be addressed (Harper, Randall, and Sharrock 2016, 15–43).

Three responses arose, and it remains important today to distinguish them. One dismisses deviations from rationalist theories as faulty and irrational behavior. The consequence is to set out and fix the behavior. "This view saves the theory and rejects the behavior" (Beach and Lipshitz 1993, 22). For obvious reasons, this reaction does not have a strong standing in psychology today—after all, the behavior did not go away, and this position does not explain why. It does, however, remain strong in studies of engineering and design behavior, where it is generally used to double down on formalization and method to bring behavior in line with theory.

> Make no mistake, in its normative role, prescribing decisions for hypothetical Economic Man, classical theory is not subject to these criticisms. It is when behavioral scientists assume that these prescriptions apply to any and all human decisions that the mischief is done. Because its prescriptive role is assumed to have the same status as in its normative role, the legitimacy of the theory is not questioned when behavior does not conform to its prescriptions. Instead, it is concluded that the behavior, and thus the decision maker, is wrong or irrational and must be made to conform to the theory. (Beach and Lipshitz 1993, 29)

The second response was to adapt rationalist theory to correspond more closely to observed behavior. This was the path taken by Nobel-prize winner Kahneman and his colleague Amos Tversky. They did not drop rational choice theory but tuned its parameters to create prospect theory (Kahneman and Tversky 1979). In the place of absolute values attributed to enumerated alternatives, they proposed that each decision-maker has a subjective value function dependent on a current state and the losses and gains *relative to* that state. It is important to recognize that they retained all other normative assumptions (Beach and Lipshitz 1993), including the implicit metaphor of the brain as a computer. Instead of abandoning these orthodox assumptions, they merely applied the axioms differently, patching up the traditional model to extend its lifespan (Gigerenzer and Selten 2001, 13–37). On this basis, they developed an influential research program on *heuristics and biases* focused on showing, empirically, how human reasoning deviates from rational standards. Ironically, the normative standard of rationality is used to show that human beings are not rational. Thomas Sturm summarizes the normative aspect of their program succinctly in pointing out that it "claims, on empirical grounds, that human beings often and systematically violate *norms of rationality* *that derive from* formal logic, probability and decision theory" (2012, 66, emphasis added). The heuristics and biases view has dominated behavioral economics for decades (Kahneman 2011; Loewenstein, Rick, and Cohen 2008), and it had a marked influence on the way that engineering and design activities are understood in computer science (Mohanani et al. 2018). Let us add some entries to the dictionary:

> **rationality**, n.: that form of deductive reasoning which can be encoded and computed.
> **irrationality**, n.: those parts of human life that rationality has no access to.

The third response arose from the recognition of methodological flaws in rationalist research.

QUESTIONABLE METHODS

The methods used by behavioral researchers in this tradition, including the illustrious pair Tversky and Kahneman, designed simplified choice situations as context-free vignettes and prompted participants to select their preferred option. It is a very efficient approach to data collection. For example, in the famous Linda experiment, they described a person as follows:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

They then prompted participants to indicate whether it was more "probable" that Linda was (a) a bank teller or (b) a bank teller active in the feminist movement. The overwhelming majority of participants elected (b). The authors interpreted this as a "massive failure of the conjunction rule" (Kahneman et al. 1982, 94), which states that statistically speaking, the compound probability of (b) can never be higher than the probability of (a). In their view, cognitive heuristics led their participants to commit logical errors. From studies like this, they drew sweeping conclusions about the supposed universal presence of cognitive mechanisms and errors.

But there's a snag, and it has two parts. First, much of this award-winning work appears to have very questionable ecological validity (Harper, Randall, and Sharrock 2016)—it describes the behavior of "people in the lab," not of people "in the wild" (G. Klein 1993, 36–50). In other words, the observed behaviors more often than not can just as well be explained (and manipulated) via the constellation of factors in the context, rather than the reasoning of the subjects themselves. For example, the effects described for the original Linda exam all but disappear with a slight reformulation (Sturm 2012; Gigerenzer 1996; Hertwig and Gigerenzer 1999, 291). Other experiments of similar structure could not be replicated at all (Harper, Randall, and Sharrock 2016). Major flaws have been pointed out in other similarly decorated studies (Harper, Randall, and Sharrock 2016, 44–81; Gigerenzer 1996; Sturm 2012).

Second, the participants' behavior can be interpreted very differently. The design of the Linda experiment started off with the mathematical theory of probability and the logical rule of conjunction. The authors operationalized these concepts loosely—very loosely—by formulating an English sentence containing the words "probable" and "and." This is a dubious operationalization of constructs. Probability theory in statistics is about repeated events, so it is not a valid theory to judge single events (Gigerenzer 1996, 593; Shafer 1986; Beach and Lipshitz 1993, 27). More generally, the logic of mathematics is not equivalent to the logic of language:

> Sound reasoning begins by investigating the content of a problem to infer what terms such as probable mean. The meaning of probable is not reducible to the conjunction rule (Hertwig & Gigerenzer, 1995). For instance, the *Oxford English Dictionary* (1971, pp. 1400–1401) lists "plausible," "having an appearance of truth," and "that may in view of present evidence be reasonably expected to happen," among others. These legitimate meanings in natural language have little if anything to do with mathematical probability. Similarly, the meaning of *and* in natural language rarely matches that of logical AND. The phrase . . . can be understood as the conditional "If Linda is a bank teller, then she is active in the feminist movement." Note that this interpretation would not concern and therefore could not violate the conjunction rule. (Gigerenzer 1996, 593)

Tversky and Kahneman appear to fall prey to an *operationalism* in which their interpretation of *probable* was defined exclusively by the mathematical operations of probability theory. This would be perfectly appropriate within the confines of probability theory, but they were operating within a broader social context in which they asked people for an opinion. Participants were evaluated on the basis of a semantic interpretation they never agreed to. And this is not simply a matter of scientific rationality trumping common sense reason. On the contrary: Tversky and Kahneman explicitly justified the former with the latter (Gigerenzer 1996, 593). By denying the relevance of content and context, they were misled into an erroneous interpretation:

> Recent studies using paraphrasing and protocols suggest that participants draw a variety of semantic inferences to make sense of the Linda problem. . . . Semantic inferences—how one infers the meaning of polysemous terms such as *probable* from the content of a sentence (or the broader context of communication) in practically no time—are extraordinarily intelligent processes. They are not reasoning fallacies. . . . Significant cognitive processes such as these will be overlooked and even misclassified as "cognitive illusions" by content-blind norms. (Gigerenzer 1996, 593)

Contradictions between rationalist theories and empirical data kept piling up, suggesting that "classical theory does not provide the conceptual depth that is needed to deal with real-world complexity; in some ways people seem far more capable than the theory" (Beach and Lipshitz 1993, 29). Harper et al. (2016, 61) conclude "we do not think the experiments these researchers undertake demonstrate these biases or aversions; in our judgment, they don't demonstrate very much at all."

There are two aspects that merit the use of rationalism as an -ism. First, note how these studies use normative assumptions as the basis for descriptive research. In linking normative theory and empirical observation, they have to face the question how to handle discrepancies between the two. We can recognize the rationalist tradition by the priority it places on theories that are normatively derived from logical principles but not empirically validated. In the face of conflict, the norms win.

Second, note how the abstraction of content and context from study design gives primacy to the normative framework's conceptual logic, while pushing situational awareness and social rationality aside. Critics advocate a broadening of the view to incorporate an awareness of the overall situation in which individuals reason and act, and to consider that, maybe, people are *reasonable* (Hertwig and Gigerenzer 1999, 300; Harper, Randall, and Sharrock 2016). I will return below in detail to the concrete implications this has on our understanding of *how system designers make decisions*, how they exercise judgment, and how we should approach studying this question. Table 7.1 briefly summarizes the mentioned rationalist studies and their

Table 7.1 Rationalist interpretations of empirical observations and alternatives to this view

	Rationalist interpretation	**Alternative interpretation**
The Linda experiment	The word "and" refers to the logical conjunction AND. Probability refers to the statistical likelihood of an event (%). The probability of the conjunction (b) must be lower than the probability of (a). Differing responses are false and reveal flaws in human reasoning. Humans use the "representativeness" heuristic in place of statistical reasoning. This leads them to commit errors. (Kahneman et al. 1982)	Probability theory is applicable to event series, not single events. The word "and" connects two parts of a sentence. Its meaning depends on the content and context. People are just as likely to interpret "probable" as plausible. The two statements are then not conjunctions. Instead, (b) indicates for many that despite her job, Linda remains committed to feminist principles. This is arguably more plausible than her having given up those ideals. Differing responses are reasonable and worth investigating because they may point to alternative views and interpretations. The study shows very little at all. "Instead of showing how people don't deploy logic, they show that they do deploy the logic of words." (Harper, Randall, and Sharrock 2016, 57)
Recent SE studies on managing technical debt	Decision-making follows rational choice, but imperfectly due to heuristics, biases, and cognitive limitations. Therefore, rational methods are better than whatever else practitioners are currently doing. When people do not follow rational methods, they behave irrationally.	When practitioners act in systems design, they exercise professional judgment in many ways. Methods are only one of many cognitive resources they rely on (Dittrich 2016; L. Suchman 2006). A plausible explanation for not using an available method is that according to the professional's judgment, the method is not suitable for the situation. This should prompt a reevaluation of the method just as much as a reevaluation of the professional's behavior. (Becker et al. 2018)

critique. It highlights that the rationalist interpretation is often questionable, that it has indeed been questioned at length, and that broadening our view beyond rationalist interpretations opens new perspectives on old questions.

THE EMERGENCE OF NATURALISTIC DECISION-MAKING

When it became undeniable that the assumptions, predictions, *and* methods of rationalist decision theory as exposed in mainstream psychology and behavioral economics were questionable, some researchers concluded that they should be set aside. This involved three shifts. First, they abandoned the "classical" rationalist *theory*: "classical theory cannot continue to be used as the standard for evaluating all decision behavior. . . . It is time to stop patching and propping an inappropriate theory. It is time to create a more useful theory" (Beach and Lipshitz 1993, 35). They started anew to develop new theories by focusing on behavior not norms (Beach and Lipshitz 1993; Zsambok and Klein 1997).

Second, they abandoned the *method* of lab experiments: "many compromises have to be made to perform controlled experiments. The restriction on context, the absence of meaningful consequences, the use of tasks with well-defined goals, and particularly the elimination of expertise in studies presenting unfamiliar tasks, all raise doubts about whether the findings of these studies can be generalized to natural settings" (Klein and Wright 2016). Instead, they went into the field and studied professionals who were making important decisions, such as firefighters, military commanders, surgeons, and engineering designers. They followed their day-to-day activities to observe and ask about the decisions they were making. This is why their research became known as naturalistic decision-making research (NDM).

Third, in contrast to the view on biases and shortcomings of human decisions, they were interested in how highly experienced individuals made decisions very well. Through this work, NDM researchers developed a toolbox of research methods and guidance called *Cognitive Task Analysis* (Crandall, Klein, and Hoffman 2006; G. Klein 2000; Schraagen, Chipman, and Shalin 2000). A central component is that they expanded the horizon of the system of interest from the individual mind to the situation and its factors. This view of *macrocognition* (Klein et al. 2003; Klein and Wright 2016) incorporates an appreciation of the relevant factors that combine in each situation differently to shape the outcomes of decision-making. It places the individuals and groups engaging in decision-making into a concrete situation and emphasizes the content and texture of that situation. This formed an entirely new paradigm distinct from rationalist research, with different base metaphors, methods, and evaluation of validity.

> Whereas the behavioral decision-making community focuses on human limitations and seeks ways to reduce biases and mistakes, the NDM community, as it performs macrocognitive research, focuses on human capabilities and regards good performance as much more than the absence of mistakes. Good performance is also about discoveries and insights; it is about the strengths of decision makers, and the importance of experience. Experience serves a variety of functions including a larger repertoire of patterns and associated actions, a richer mental model of how things work to support inferential reasoning and sensemaking for diagnosis and anticipation. (Klein and Wright 2016, 3)

What they found was remarkable (G. Klein 1998; Zsambok and Klein 1997). Initially, they found no "decisions" at all: people denied making choices among options. This made little sense at first. It was only when they broadened their search beyond "choice" that *decision-making* appeared: people did commit to actions, but not by evaluating options against criteria. Importantly, even professionals trained in MCDA, and who believe in its value, don't *use* it much (Isenberg 1984). Instead, in recognition-primed decision-making, a prominent model that has frequently been documented, people rapidly identify and process cues in the environment that allow them to match salient features of a situation against their experience, and they use their experience and technical knowledge to generate one plausible course of action. They then use various techniques, above all mental simulation, to evaluate how well this option would work. If they are satisfied, they proceed; if not, they adapt the option or drop it to generate a new path of action. They never perform pairwise comparison, and they have no need to explicitly articulate criteria for evaluation (G. Klein 1997).

DEFENSES OF RATIONALIST THEORIES

Two explicit defenses for the rationalist paradigm are typically brought forward against its critiques. One defends the rational process as a regulative ideal worth striving for (Baron 2012), even if we "fake it" in design (Parnas and Clements 1986). This defense has normative validity—it is still often useful and valuable to justify decisions in terms of systematically specified criteria and to articulate which alternative options were considered. (I did a PhD on that and have few regrets.) But this argument does nothing to justify the attribution of *descriptive* validity. Normative value does not beget descriptive validity (Beach and Lipshitz 1993). There *is* to a limited degree a *performativity* in normative models: for example, to some degree normative economic models *produce* the behavior they describe (MacKenzie and Millo 2003). But this aggregate observation does not imply individual behavior and is certainly not valid for engineering and design methods (Dittrich 2016). It is naïve to assume that teaching rationalist methods will simply produce behavior that corresponds to them. In addition, this defense fails to address the deeper concerns about the issues raised by the reified, operationalist frame of thought that is introduced to the discussion and study of human behavior by the metaphor of thought as information processing. Methods are not programs to be run on the computing hardware of team member's minds, and computing education does not install an operating system in students' heads.

The other defense asks, "what's the alternative?," following a long-standing tradition that relegates anything outside narrow rational behavior into the realm of irrational behavior (Erickson et al. 2013; Sturm 2012). It is important to recognize that this argument is operationalist itself: it fails to recognize anything not defined by its own specification as a valid form of reasoning and leads, ironically, to a false dichotomy that considers the opposite of rationalist to be irrational behavior.

This is not to deny that people are also perfectly capable of using rationalist methods as part of their cognitive toolbox. These methods are simply not always appropriate, and the heuristics people use instead often outperform rationalist methods (Gigerenzer and Selten 2001). Unsurprisingly, it proved difficult to convince the rationalist camp to change course, despite conciliatory voices emphasizing commonalities across the schools of thought (Kahneman and Klein 2009). The relationship between descriptive and prescriptive research on decision-making is complicated (Bell, Raiffa, and Tversky 1989) and remains so. Advances in neuroscience suggest that multiple modes and systems of decision-making seem to coexist in the brain. Rational models do get used for clearly circumscribed tasks, while methods such as recognition-primed decision-making and other strategies identified by naturalistic research are used in many less circumscribed situations, for example to structure problems (Loewenstein, Rick, and Cohen 2008; Kahneman and Klein 2009).

Systems design research needs to take both views and understand their relationships. We need to question whether the boundary of decision-making supposed by rationalist theories are meaningful and helpful for us in understanding design for sustainability and justice, or whether it would not be smart to shift our focus: from choice to judgment, from rationalist theories to naturalist theories, and from the mind as a computer to the social situation of decision-making.

BEWARE THE NORMATIVE FALLACY

Having explored how other fields struggled to extricate themselves from the grip of an obsession with a limited theory, let us take another close look at the myth of rational decision-making (RDM). RDM manifests as myth when its scope of relevance is overextended. In computing research, RDM has widely been uncritically accepted as "the" theory of decision-making, and the nuances explored above often collapse into one "decision-making" concept that is defined as choice between enumerated options.

What are the consequences? Here I will explore one in more depth: RDM has socially preformed empirical software engineering (SE) research and misled it to severely misinterpret how people use methods, make judgments, and arrive at decisions. For good reasons, MCDA is central to SE methods for making tradeoff decisions: It offers a rigorous, systematic, repeatable and teachable approach to making decisions based on solid evidence, and it facilitates the review of that evidence. For decision situations that meet specific criteria, there is little doubt that these methods should be applied correctly. But how people actually use these methods, and what it means to "use a method," differs from what RDM assumes, and that places empirical research on practice in tension with the prescriptive nature of method development (Dittrich 2016). As mentioned earlier, the rationalist tradition sees methods as programs, and this is unsurprising given the role of concept-mapping. When computer science curricula focus almost completely on how programs work, it seems natural to rely on explanatory schemas from the domain of programming to understand what *programmers* do. To those captive to RDM, any deviations appear as irrational mistakes. In this clash between theory and practice, practice loses, but so does theory. “Divergence from what method developers and research community recommend and diversity of practices should not by default be regarded as a problem but as source of understanding the rationalities of practice that then can inform method development and appropriation” (Dittrich 2016, 228).

The orthodox view of “methods as programs” is based on an operationalism that thinks of the elements of design and development activity as clearly definable discrete atomic operations linked by specified relationships. Process modeling formalizes these elements and relationships. A method is defined by these operations, and deviations from the operations in practice are either invisible or seen as defects. But this operationalism does not merely mean that everything that matters can be operationalized—it also defines, reversely, what gets recognized as belonging to a class of things. The operationalism often structures entire research projects, preconfiguring behavioral studies of engineering activity to only recognize those parts of practice that fit the decision-making operations defined by MCDA.

Behavioral studies of systems design matter because understanding how design decisions go right or wrong is a key step to doing better. But the operationalist understanding of RDM leads many studies of engineering and design practice to commit a normative fallacy (Campbell 1970)—they misappropriate normative theories and prescriptive models for descriptive purposes.

Suppose we study the behavior of participants facing a risky software project situation. A series of twenty bugs were identified in short sequence in a new system under development. They appear somehow related, but it is unclear how. One is critical, and nineteen are severe. The team considers the business value of fixing one critical bug equivalent to fixing four severe bugs. Most team members want to focus all attention on the critical bug first (strategy a), but one claims to know how to fix all twenty and wants them to pursue their strategy (b). Our study participants must choose between two competing strategies: In their judgment, strategy (a) is almost certain to resolve the critical bug in the system, strategy (b) has a decent shot at fixing all bugs.[8] Many participants readily choose (a).

According to the normative model, this is, strictly speaking, an "error" because in its calculation, the expected value of option (b) is much higher. For example, if we set (a) at 90 percent and (b) at 33 percent, we arrive at expected values of 0.9 against 1.9. A sensitivity analysis may confirm that the two options' ranks are very robust against estimation errors.[9]

The question is not whether we agree with this assessment. The normative fallacy comes into play when we frame research questions and collect data *about* this situation. If we set out to describe how participants make their choice, should we ask them "how did you compute the value of each option?" and "how did you weigh your factors?" Doing so would mean committing the normative fallacy: It would assume that they decide by applying numeric operations to abstract concepts without further considering the content and context of the situation they are in to reflect on the meaning of these numbers. Instead, we could ask an open-ended question: "How did you make your choice?" If given a chance, participants may rightfully introduce concerns outside the framing of the gamble which make it perfectly reasonable for them to prefer option (a). They might say that it can be communicated more effectively to the stakeholders, hope that fixing the critical bug produces insights that help fixing the others later, or believe that resolving the critical bug first will reduce the stress on the team. They may also emphasize the nature of ambiguity: After all, the percentage estimates are in truth based on ambiguous information. Their concerns and judgments transcend the original framing of the gamble and situate it in a broader context. They seem much more reasonable than the normative assumptions of rational choice, but they do not fit into the confines of "rational decision-making."

THE NORMATIVE FALLACY IS COMMON

In behavioral SE studies, unfortunately, this normative fallacy is common. Our systematic literature reviews in key areas found that many studies use prescriptive theories to collect and interpret data about behavior for descriptive and explanatory purposes (Becker et al. 2018; Becker, Walker, and McCord 2017). Most importantly, the normative fallacy manifests as studies in which data collection and analysis are incorrectly predicated on the narrowly framed concepts of factors, weights, ranking, and choice.

Broader aspects of reasoning including the role of expertise, experience, cognition, incentives, mental simulation, judgment, and perception are never considered. For example, one study about technical debt decisions asked participants these questions:

1. What factors are considered when you make a decision about when to fix a defect?
2. How are these factors weighted? (Snipes et al. 2012)

Responses to these questions were taken at face value. This may appear normal, and it is normal in SE research. But accepting the findings at face values means accepting that the actual decisions of these participants are *adequately described* by the operations specified in MCDA: criteria specification, weighting, evaluation, ranking, and choice. That is, these questions make no sense unless we take the normative theories of rational choice as descriptively valid. That is a mistake, so the questions above simply failed to capture how the participants really made their decisions. Instead, they nudged them to describe their reasoning *as if* they had considered weighted factors.

Even though most participants in these studies deviate from the proposed norms, the norms and assumptions are never questioned. Instead, the deviations are considered deficiencies: errors that need to be fixed. For example, in another study in which the observed practices did not correspond to the proposed method, the researchers developed a data categorization scheme predicated on MCDA to make sense of it. We are left to wonder what they would have found had they looked beyond it. Similar bias pervades many discussions in which numerical optimization across multiple objectives is depicted as automatically superior to individual expertise and team knowledge, as in this case: "Technical debt . . . is currently managed in an implicit way, if at all. Decisions are largely based on a manager's experience, or even gut feeling, rather than hard data gathered through proper measurement" (Guo, Spínola, and Seaman 2016, 160). But what is called *gut feeling* here has long been relocated to the brain and reinstated as social rationality, experience, and judgment.

This is not to say that participants in this example *should not* consider a set of factors and gather evidence: simply that their reasoning will remain invisible to those who perform the study. Crucially however, because these questions are posed in the context of a scientific study by an academic research team with scientific credentials, the participants certainly *will* provide answers. In doing so, they will retroactively construct plausible factors and weights. Papers that uncritically report these answers, though not empirically valid, often pass peer review because reviewers in SE are also not trained in behavioral research or psychology. The findings are then cited to support further research, most of which is again prescriptive. In this way, misunderstood empiricism reinforces a misleading narrative. The myth of rational decision-making is reproduced via its retelling.

Imagine if the study above had instead asked an open-ended question: "How did you decide when to fix a defect?" This simple reorientation would already avoid two important misunderstandings produced by the torque of the RDM myth. First, it sets up the expectation that the participants' reasoning will be described on their own terms and leaves open in what way the result will be compared and contrasted with various theories of decision-making. Second, when and if such comparisons happen, if the collected responses indicate a divergence between rationalist models and participants responses, the open-ended structure of their responses will make it easier to see when their reasoning is different. It will make it easier to recognize that their decision-making is not an impoverished *version of* MCDA but could instead be interpreted as balancing competing priorities in a complex situation. In other words, we should get ready to entertain not only the possibility that our theory is wrong but question how it prestructures our investigations. Table 7.2 summarizes this

Table 7.2 The normative fallacy in empirical research paradigms

		Normative research (what should be?)	**Empirical research (what is?)**
Theoretical framework	Descriptive and explanatory	**OK**: normative recommendations grounded in empirical findings	**OK**: empirical research
	Prescriptive	**OK**: deductive or constructive research oriented towards design or modeling	**NOT OK**: normative fallacy[a]

Note: a. Testing theoretical assumptions of prescriptive theories *as hypotheses* in empirical research is of course fine, but that is not a case of this quadrant: In this case, the theory takes the role of research object, not of theoretical framework.

discussion by illustrating, in simplified form as "ideal types," the possible combinations that arise from the use of descriptive or prescriptive theories in normative or empirical research.[10]

The normative fallacy carries important consequences:

1. Participants *will* respond to the questions above, and those responses will reinforce the empirical basis for normative theory, even when this empiricism is flawed and tautological.
2. Because of the unquestioned acceptance of RDM as normative *and* descriptive standard, any deviation will be treated as a defect and labeled as bias or error.
3. Participants' rejection of decision-making methods is often explained away with the naïve suggestion that they just need better training or more time so they will see the right way.
4. In all this, the uncritical empiricism built on the normative fallacy ultimately turns into ideology. And it keeps us from asking the important questions: How *do* people make tradeoff decisions when the outcomes are at a distance? What can we learn from those who act with long-term vision? Do our methods inadvertently tilt the field against sustainability and justice? How can we do better?

CONCLUSIONS

The myth of rational decision-making tells us a story of criteria, options, weights, ranking, and choice. This is not what happens, even for those people who are trained in and believe in these methods. In the end, it appears that "humans are not two eyeballs attached by stalks to a brain computer" (D'Ignazio and Klein 2020, 85). The rationalist logic of choice is one tool in the toolbox of human reasoning. When applied as a description, it is narrow, flawed, and misleading. People are perfectly capable of deploying rationalist models and methods, but they often choose not to, and they often have good reasons.

It is hard to get over the myth, because the reluctance to abandon cherished paradigms is always strong (Kuhn 1962; Ralph and Oates 2018; Ralph 2018). When theory and data collide, "the possibility that the theory is inappropriate is seldom entertained" (Beach and Lipshitz 1993, 28). The unquestioned adoption of rationalist theory unfortunately continues in some computing fields despite the indisputable fact that the rationalist model of decision-making misses countless ways in which smart humans reason intelligently and function effectively in complex open-ended environments like systems design. It misses situated and experiential knowledges, moral reasoning, judgment, and the nuanced nature of intuition, recognition, and expertise (G. Klein 1998). How is a theory of choice based on the human mind as a defective calculating machine supposed to lead us into a future in which everyone designs responsibly, sustainably, and justly? It won't.

But if that is not how it happens, what else is happening? How do people take all these decisions in systems design that affect, at a distance, those in the future and far away? Without understanding that, we cannot possibly hope to design more sustainably, and more justly.

Moving beyond the myth of rational decision-making, and avoiding the normative fallacy, opens new opportunities. We can reorient ourselves: from methods as program to methods as resources in situated action; from people as information processors to people as reasonable and purposeful social beings; from engineering practice as applied science to engineering practice as a social epistemic practice on its own ground; from rational choice to reasonable decisions and wise judgments; and from rationalist quasi-experiments to naturalistic studies. We can focus on identifying wise judgments and supporting reasonable decisions based on macro-cognitive perspectives. We can study judgment *and* decision-making accounting for reflection, critical awareness, the critique of boundaries and value systems, and the composition of the macro-cognitive environment in a decision situation. We can aim to understand which constellations are more likely to produce short-sighted decisions, and we can support design teams in redesigning the architecture of their situations, so they are better able to take a long-term view and consider distant impacts and outcomes of their decisions. Chapter 11 explores this path.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Becker, Christoph (Director of the Digital Curation Institute), author.
Title: Insolvent : how to reorient computing for just sustainability / Christoph Becker.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022038283 (print) | LCCN 2022038284 (ebook) | ISBN 9780262545600 | ISBN 9780262374651 (epub) | ISBN 9780262374668 (pdf)
Subjects: LCSH: Electronic data processing—Social aspects. | Computer systems—Environmental aspects. | Information technology—Social aspects. | Sustainable development.
Classification: LCC QA76.9.C66 B435 2023 (print) | LCC QA76.9.C66 (ebook) | DDC 303.48/34—dc23/eng/20221121
LC record available at https://lccn.loc.gov/2022038283
LC ebook record available at https://lccn.loc.gov/2022038284