A Consequentialist Model of Simplified Reasoning

Sebastián Sánchez-Martínez[1]

Workshop for Theoretical Philosophy, TU Dresden
02 February 2024

[1] mail: sesanchezma@pm.me
web: www.sesanchezma.com

Thanks for your feedback!

A typical way of seeing simplification is by asking whether, at the end of it, we were right. But we also seem to think about it in terms of the process. So, something about the strategy of simplification gets unnoticed when we focus merely on the end result. A focus on the process can be beneficial to get clear about our target of evaluation. My proposal: make sense of this line of argument by offering a consequentialist model. Under this new light, simplification is pictured as a strategy for strategy-selection, and one that best responds to certain limitations.

Spoiler:This picture can be puzzling to epistemologists, because it seems to leave the role evidence floating in the air and doesn't seem to permit, under limitations, to know things we are in a position to know.

Too Abstract! Let's...

Plan:
1. Make it Simple
2. Harman on Reasoning
3. Simplification, Complicated
4. A Consequentialist Model

Make it Simple

I will motivate my approach using two examples that highlight the use of simplification in both practical and theoretical reasoning.

September heat: *Madeleine is in Madrid during the September heat wave. She reads in the newspaper that, given the historical record, it is less than 0.05 likely that this September is going to rain. On this basis, in planning her day today in Madrid, she formed the intention to go out without an umbrella. Has she made her no-umbrella decision using her credences in rain together with expected utility maximization? No, she is a simplified reasoner and has just taken for granted that it is not going rain in Madrid during the September heat wave.*

Baseball Game: *"suppose Rainer knows that the LA Dodgers will be playing against the Seattle Mariners, and he wants to determine the probability of the following proposition (call it g): that it will rain during the game. Suppose, however, that Rainer doesn't know which team will have home field advantage. In this case, he might arrive at the probability of g by taking the weighted average of the probability of g conditional on the game being in LA and the probability of g conditional on the game being in Seattle. In so doing, however, he would be treating as true, in the sense just defined, the proposition that the game will be held in either LA or Seattle. And yet, if he is rational and has ordinary evidence, then he will have nonzero credence in a multitude of alternative possibilities."*[2]

I am going to take these examples as representative of the ones used in recent philosophical literature to motivate a discussion about simplified reasoning.

Two possible approaches to simplified reasoning: attitude-centered and process-centered.

[2] *"an agent treats a given proposition p as true just in case she evaluates her alternatives by the same procedure by which she would evaluate them conditional on p." See: [Ross and Schroeder, 2014, p. 266]*

Attitude-centered approach to SR: These cases are mostly used to ask questions about which doxastic attitudes bear some badges of epistemological interest.

For instance, we could say something about what, if anything, gives evidential support to Madeleine to take for granted that it's not going to rain, despite the possibility of rain.

And we could equally say something about what gives justification to Rainer to take for granted that the game will be either in LA or Seattle, despite the 'multitude alternative possibilities'.

The leading motivation behind asking those kinds of questions is that there are different *reasoning types* that produce in their own way certain attitudes that can be normatively assessed.

i. e.: simplified versus complex types.

On this approach, one would take Madeleine's and Rainer's attitude of *taking for granted* and assess it for their for its rationality. Some other notions can be used as standards of that assessment: truth, evidential support, accuracy, etc.[3]

[3] Accuracy is a technical term: accuracy: credences :: truth : beliefs. See: [Joyce, 1998]

I'm *not* going to argue: what, if anything, makes it rational for Madeleine or Rainer to adopt attitudes of taking for granted or taking as true?

I am *also not* arguing for why the attitude-centered approach is wrong.

I don't think (yet) is wrong.

I am just stating it is a different project wrt the following:

Process-centered approach to SR: talking about reasoning types in terms of simplicity versus complexity $\longrightarrow$ another approach to simplified reasoning.

This other approach would make salient that to each reasoning type corresponds a different *cognitive process*. The idea is that both Madeleine and Rainer could have reasoned *probabilistically*, but have simplified their reasoning by not doing so.

This means just that what Madeleine intends to do is based on a an all-out attitude, though she could have applied a utility maximization rule and used credences. Analogously: Rainer's credence in rain is based on an all-out attitude that the game will be either in LA or Seattle.

In probabilistic reasoning, taken as a cognitive process, the transition to the end states of reasoning is made using graded attitudes, as opposed to all-out ones. So: SR *qua* process is not probabilistic.

Question 1: Are cognitive processes (corresponding to different reasoning types) bearers of badges of epistemological interest?

Common answer: Yes.*

*But...

The common answer by attitude-centered authors relies on a contrast between *evidential* and *practical* normative standards or between *evidential* and *all things considered* normative standards. [4]

[4] Depending on who you ask, epistemic = evidential can be controversial.
But for my purposes we could replace 'evidential' with 'reliabilist' or 'proper functioning' or 'epistemically virtuous'.

The contrast is applied in different normative assessments. One would thus say that there is a difference between having *evidential reasons* to take for granted that it's not going to rain and having *practical reasons* to do so.

Purely evidential requirements don't to rule out that there is going to rain, so one could say that evidential certainty doesn't rationalize taking for granted that it's not.

But practical requirements would reasonably allow for a separable

notion of practical certainty that does seem rationalize that taking for granted.

Something similar would give Rainer's *practical* reason to treat *as true* the disjunction $(LA \vee Seattle)$, despite the mentioned nonzero credence in multitude alternatives.

Separating this way notions of evidential and practical certainty, could lead to different moves. These range from saying that practical considerations have the effect of changing evidential requirements (pragmatic encroachment) or that practical considerations don't have such an effect on *beliefs* or *credences* but on other attitudes, such as *acceptance* [Bratman, 1992] and *degreed acceptance*. [Dinges, 2022]

Similar move: morally supported attitudes that lack evidential support or attitudes with good evidential support that are morally unacceptable.

Also of interest for the attitudes-centered theorist: Which attitude best suited for the job required by a reasoning type? See: [Staffel, 2019, Dinges, 2021, Palmira, 2023] **Let's talk about details in the Q&A**

In a nutshell: rationality of the process $\longrightarrow$ rationality of the attitude.

Not explicitly presupposed, but implied: normative evaluation of a cognitive process depends on whether the reasoning outcome (belief, credence, acceptance) is bearer of an epistemological badge (evidentially supported or not, rational or not, etc.)

Substitution: Is the cognitive process of SR rational? $\longrightarrow$ Is the outcome of SR rational?

Answer to the substitute question: SR can be rational once we separate an epistemologically interesting notion of rationality and apply it to the reasoning outcome. Typically: propositional rationality and doxastic rationality [Firth, 1978]

Holmes studies the mud on the wheels of the carriage.
Outcome: the coachman did it.
Propositional rationality = bears on the logical part of belief (propositional content)
Doxastic rationality = bears on the attitude, requires rational basing on the evidence.

Problem: this answer is indirect!

Question 2: Can we answer directly to Q1?

My answer: Yes. Helpful: cover more ground before tackling the process-centered project.

Two keys: First, a more precise process-centered model of reasoning. Second, extending that model to SR.

Harman on Reasoning as Change in View

The broad model of reasoning will be *linear*.[5]

[5] [Harman, 1986] Chapter 1.

A more fine-brushed model is Johnson Laird's mental model theory. Reasoning = iconic representation of possibilities. See: [Johnson-Laird, 2008]. Also for contrast see [Mercier and Sperber, 2017].

Reasoning is represented as a temporally extended sequence of operations resulting in a *changed view*. A 'view' can be understood as a set of practical and doxastic attitudes. For instance. V_{t_0} = {intention to come to TUD, belief that the tram will be working, belief that I will come to TUD by tram}.

Defined this way, a 'reasoned change in view' could be described in terms of adding or subtracting new attitudes.

For instance, in light of seeing that the tram is not working I can give up my belief that the tram will be working and either maintain or abandon the intention to come to TUD.

In maintaining my intention that I will come to TUD, I further change my view to V_{t_5} = {intention to come to TUD, belief that the tram will not be be working, belief that I will come to TUD by bike}. Figure 1 can be a broad representation of this change in view.

In the examples of Madeleine and Rainer, as well as in this

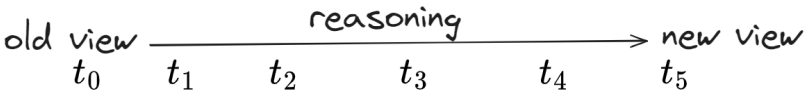

Figure 1: This is a rough initial representation of a reasoned change in view.

new one, I appeal to the different attitudes that end up changing in reasoning.

Some of the changes that occur correspond to theoretical attitudes and some to practical ones. These attitudes may change simultaneously, but on this model this is best captured as different segments each corresponding to the theoretical and the practical attitudes.

GOING BACK TO THE INITIAL EXAMPLES, Madeleine and Rainer received information that allowed them to form *credences*. Having reasoned with all-out attitudes, this seems to involve a *selection* of a non-probabilistic mode of reasoning.

Since they begun at t_0 with evidence supporting some credences, there is a sense in which both started reasoning with credences.

But, crucially, it seems like both of them took an optional all-out attitude:

- Madeleine: it was possible to reason using $c(\text{rain in Madrid}) = 0.05$, but she reasoned instead by taking for granted that it will not rain in Madrid.
- Rainer: it was possible to reason using $c(LA \vee Seattle) < 1$, but instead he reasoned instead by taking for granted that $(LA \vee Seattle)$.

Thus, having taken the simplified strategy involves the existence of an alternative one. For this reason, it can be helpful to think about both of them *selecting* reasoning without the credences.

Assumption: reasoning can be done without all-out or degreed ones or with a combination of both. Reject: Harman's view about the psychological impossibility of probabilistic reasoning. See [Staffel, 2013] for discussion. **Let's talk about details in Q&A.**

Further help: distinguish between two levels of cognitive operation: strategies and steps of reasoning.

- Reasoning strategy = optional sequence of reasoning steps taken with an end-goal in mind. Thus, a strategy is one of the alternative reasoning lines that a reasoner can use to solve a particular problem.
- Reasoning steps = the particular algorithm used to solve a reasoning problem. These steps could be thought as each one giving an instruction to the reasoner to perform a certain operation, like adding, subtracting, eliminating, etc.

Leads to a refinement of the model of reasoning. We need to allow for diverging lines, where each alternative line corresponds to one of the reasoning types (see Figure 2):

In sum: a linear model of reasoning like Harman's allows to represent the temporally-extended steps that lead from one set of attitudes to another. To apply it to our examples of SR, the model needs to be extended to include a choice-point where each strategy (S_C and S_S) corresponds to different sequences of steps.

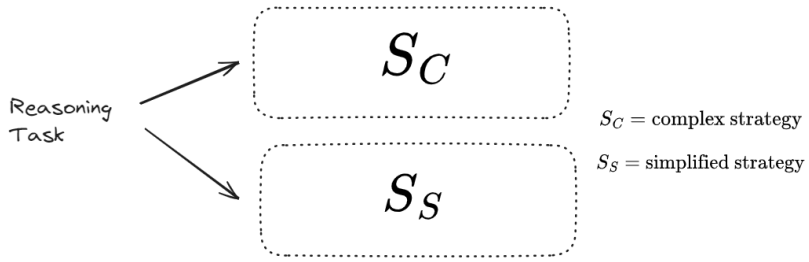

Figure 2: Complex and simple alternatives

- Madeleine's S_C could be: credences plugged in in a utility maximization rule.
- Rainer's S_C could be: assigning a credence < 1 to the game being in $(LA \vee Seattle)$. See Table 1 for details. [6]

Since strategies are defined as goal-dependent, both the complex and the simplified reasoning types correspond to different goals that our agents have in mind.

Goals could be deciding whether or not go out with an umbrella (Madeleine) or getting unconditional probability of rain (Rainer).

This further suggests to think about alternatives in reasoning using notions of means-end rationality. Harman is explicit in endorsing this way of thinking about reasoning in goal-dependent ways.

[6] The more alternative possibilities, the lower the credence in $(LA \vee Seattle)$, and the less confident in $c(g)$ he will be, maintaining constant values of $c(g|LA)$, $c(g|Seattle)$, and the weighted average rule.

$c(LA \vee Seattle)$	$c(g)$
1	0.55
0.99	0.5445
0.98	0.539
0.97	0.5335
0.96	0.528

Table 1: Assuming constant values of $c(g|LA) = 0.7$ and $c(g|Seattle) = 0.4$, and further assuming constant the rule of weighted average, lowering the credence in $c(LA \vee Seattle)$ will result in a lower credence to the unconditional credence in rain, $c(g)$.

I hypothesize that many of one's decisions are of necessity what we might call simple decisions. These arise when one finds oneself with a salient end E and one recognizes a salient means M that will get one E. In a simple case, one does not consider whether there might be some other means to E or some other end distinct from E that one might now obtain, and one disregards any other consequences of one's act. One simply forms the intention of getting E by doing M. [Harman, 1986, p. 106]

[...] in a simple case one forms the intention of getting E by doing M without thinking about side effects. If one happens to notice side effects or consequences, that does not by itself normally influence one's decision. All this is so unless there is something that sets off an "alarm" and even then, one merely checks to see whether the alarming consideration is sufficient to lead one not to make the decision one would otherwise make. [Harman, 1986, p. 106-7]

Key Takeaways:

- First, the difference between what the evidence supports and actual reactions in cases of SR suggests the existence of alternatives.

Sometimes, what the evidence supports doesn't correspond to the outcomes from how one ought to simplify reasoning.

- Second, in order to model simplification as the selection of an alternative, it's not necessary that selection is deliberate and explicit.
- Third, the existence of alternatives gives support to similarities between theoretical and practical reasoning.

These three takeaways suggest a connection to the use of decision-theoretic tools to model theoretical reasoning can bring cases of SR into a single model. [7]

[7] Classical example of decision-theoretic tools applied to model forecasting, estimation and generalization from observation: [Levi, 1967]; recent application of the model to the cognitive act of guessing: [Dorst and Mandelkern, 2021].

Simplification, Complicated

BUILDING ON THE PREVIOUS MODEL ...

Question 3: What drives simplification?

A hunch: computational limitations. Sometimes, more time is too much, because it's not expected to make a valuable difference. [8]

[8] More serious support for the hunch: Early work: see [Simon, 1978, Simon, 2008]. More recent approaches: [Lieder and Griffiths, 2020].

Think of the main difference between complex and simplified strategies as a **difference in time**. The leading idea is that, all things being equal, complex strategies require more time than simplified ones.

If simplicity means, all things being equal, less time ... we can expect:

- **Goal-Sensitivity:** reaching the outcome of reasoning is right wrt to true results about something uncertain at the outset.

  Not to carry the umbrella: did it rain or not?

  Credence in rain, $c(g)$: where is the match going to take place?

  Notice: what is taken for granted is function of the question. And: the question allows for wiggle room.

  Upshot: more time reasoning guarantees more accuracy, but there's no correspondent guaranteed difference in answering the questions. Ignoring small differences will be inconsequential, while considering them will be felt as cognitive effort.

- **Deadline-Sensitivity:** simpler strategies allow to finish reasoning earlier. Usually, reasoning to settle a question is merely a part of a longer train of thought. Sensibly, other things in planning her day in Madrid will need the same reasoning resources (the same time) Madeleine is using to settle the umbrella question. Other things being equal, finishing earlier can be expected to bring better results in the context of other reasoning problems. Upshot: at certain contexts, more time in reasoning can be detrimental.

- **Pragmatic-Sensitivity:** goal sensitivity: simpler strategies get good enough results. Deadline-sensitivity: finishing earlier is better when other tasks compete for the same time. But sometimes the *ceteris paribus* clause is false. Time consuming considerations bring other non-reasoning things to bear on the value of reasoning.

  Going for a walk: spending more time using expected utility maximization eats up walking time for Madeleine.

  Being hungry: causes delays in reasoning when it's simple, so more delays for the complex. This can reasonably bias a preference for the simple strategies.

  Financial costs: expected big stakes have an impact too. Familiar from the literature on pragmatic encroachment.

A Consequentialist Model

Assumption in the literature on SR: it is a strategy common to both practical and theoretical reasoning.

Intuitive considerations: cases of SR seem to conform to familiar cases of everyday reasoning, and examples have a prima facie flavor of verisimilitude. Parallel is clear in cases in the literature: [Harsanyi, 1985, Bratman, 1992, Lance, 1995], etc.

Theoretical considerations: ignoring small error possibilities seems key to generalize and put limits on ideal models of practical and theoretical reasoning. One reason is that these seem to bee too demanding for limited reasoners; another is that human reasoners don't seem to conform to these demands. There has been recent push to develop 'non-ideal' theories that bridge the resulting gap, which comes from assuming key commonalities between theoretical and practical reasoning.

While there is controversy about what the mechanics of simplification could be and what to make of the empirical evidence suggesting departure of cognitive models, there seems to be agreement that a unified model of SR for both theoretical and practical reasoning could illuminate the important commonalities.

There are plausible motivations for a specifically consequentialist model:

First, there is empirical literature supporting that humans are surprisingly good at coping with highly demanding and uncertain environments. Also, there is evidence that heuristics may be rationally selected along the lines of a cost-benefit meta-analysis.

Second, a well-nurtured list of cognitive biases are on track of being vindicated by pointing to rational models that produce them.[9] This is obviously controversial and too recent to predict how convergence might come, but it does provide a reason to think a consequentialist model is a candidate for thinking about cases of simplification.

[9] [Kelly, 2008, Polonioli, 2014, Morton, 2017, Icard and Goodman, 2015, Hedden, 2019, Dorst, ming, Thorstad, 2023]

Third, my proposed generalizations (sensitivity to goals, deadlines and pragmatic factors) of SR point towards patterns of criticizing agents failing to simplify when complexity doesn't make a difference.

The model will be made out of familiar ingredients in means-end rational analyses:

In solving a reasoning task we are faced, in the simplest case, with a choice-point between a set of strategies $\mathcal{S}$ encompassing a simple strategy S_S and a complex one S_C.

Strategy selection is thus a function taking the set of strategies and their consequences. Importantly, in this model consequences taken into account are given a measure of time. More time in cases of need for more reasoning, but also more time in cases where a simple strategy will not suffice, given the features of the reasoning task.

The result is a maximization strategy of the value of pursuing the reasoning goal.

See [Lieder and Griffiths, 2020, p.5] and [Icard, 2023, pp.80-107] for more details about how to formalize the model.

I will assume that reasoners are unsure of both the consequences

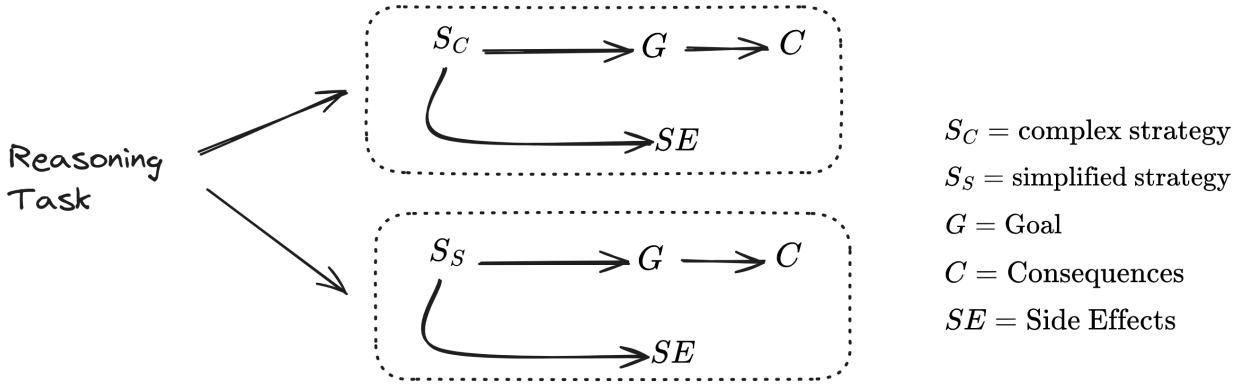

Figure 3: Reasoning with a preview of possible consequences and side-effects. The proximal goal of reasoning would help define how one's reasoning is instrumental in getting the reasoning's goal

and the side-effects of each one, but that they estimate them from the features of the problem based on previous learning or hardwired tendencies. And the value of reasoning comes from making an approximate estimate of the payoff and the costs in terms of time of the reasoning strategy.

Objection: How do we know that strategy selection stops at this level of metareasoning and doesn't continue at other higher-order levels?

Two sources of reassurance:

We can specify conditions under which, over time, we learn how do distinguish environments where strategies work and where they don't.[10] For instance, heuristics like take-the-best work better in non-compensatory environments and weighted-average rules work better in compensatory environments.

[10] [Russell and Wefald, 1991, Rieskamp and Otto, 2006, Vul et al., 2014, Lieder and Griffiths, 2015]

In addition, learning which strategies work best need not be a matter of individual reasoners solving problems from scratch. Systems of literacy and education actually give us a lot of tools to cope with problem-solving. And evolution migh have selected for a lot of reasoning instincts or hardwired tendencies under certain circumstances.

Wrapping Up

Thinking about SR as a reasoning type requires a working model of reasoning. A simple model can be Harman's model of reasoning as change in view.

A consequentialist model emphasizes important similarities of theoretical and practical reasoning. When you are deciding what to do you are faced with options, which is similar when you have different options about what you should believe or how you should set your credences.

Evidential consideratons are not exclusively the ones that need to be taken into consideration. The model does construes the rationality of SR as an all things considered matter.

But this can be puzzling to epistemologists, because sometimes

simplified reasoning make us leave evidence on the table and prohibit us from knowing things we are in a position to know.

Next stop: connection with the literature on inquiry, which seems to have a similar process-centered approach to epistemological issues. Think about instrumental principles of inquiry on [Friedman, 2020] and more generally on the project for a zetetic epistemology on [Friedman, ming].

References

[Bratman, 1992] Bratman, M. E. (1992). Practical Reasoning and Acceptance in a Context. *Mind*, 101(401):1–16.

[Dinges, 2021] Dinges, A. (2021). Beliefs Don'T Simplify Our Reasoning, Credences Do. *Analysis*, 81(2):199–207.

[Dinges, 2022] Dinges, A. (2022). Degrees of Acceptance. *The Philosophical Quarterly*, 72(3):578–594.

[Dorst, ming] Dorst, K. (forthcoming). Rational Polarization. *The Philosophical Review*.

[Dorst and Mandelkern, 2021] Dorst, K. and Mandelkern, M. (2021). Good Guesses. *Philosophy and Phenomenological Research*, page phpr.12831.

[Firth, 1978] Firth, R. (1978). Are Epistemic Concepts Reducible to Ethical Concepts? In Goldman, A. I. and Kim, J., editors, *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*, Philosophical Studies Series in Philosophy, pages 215–229. Springer Netherlands, Dordrecht.

[Friedman, 2020] Friedman, J. (2020). The Epistemic and the Zetetic. *The Philosophical Review*, 129(4):501–536.

[Friedman, ming] Friedman, J. (forthcoming). Zetetic Epistemology. In Flowerree, A. K. and Reed, B., editors, *Towards an Expansive Epistemology: Norms, Action, and the Social Sphere*. Routledge.

[Harman, 1986] Harman, G. (1986). *Change in View: Principles of Reasoning*. MIT Press, Cambridge, Mass.

[Harsanyi, 1985] Harsanyi, J. C. (1985). Acceptance of Empirical Statements: A Bayesian Theory Without Cognitive Utilities. *Theory and Decision*, 18(1):1–30.

[Hedden, 2019] Hedden, B. (2019). Hindsight bias is not a bias. *Analysis*, 79(1):43–52.

[Icard, 2023] Icard, T. (2023). Resource Rationality.

[Icard and Goodman, 2015] Icard, T. F. and Goodman, N. (2015). A Resource-Rational Approach to the Causal Frame Problem. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*.

[Johnson-Laird, 2008] Johnson-Laird, P. N. (2008). Mental Models and Deductive Reasoning. In Adler, J. E. and Rips, L. J., editors, *Reasoning*, pages 206–222. Cambridge University Press, 1 edition.

[Joyce, 1998] Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65(4):575–603.

[Kelly, 2008] Kelly, T. (2008). Disagreement, Dogmatism, and Belief Polarization. *Journal of Philosophy*, 105(10):611–633.

[Lance, 1995] Lance, M. N. (1995). Subjective Probability and Acceptance. *Philosophical Studies*, 77(1):147–179.

[Levi, 1967] Levi, I. (1967). *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge: MIT Press.

[Lieder and Griffiths, 2015] Lieder, F. and Griffiths, T. (2015). When to use which heuristic: A rational solution to the strategy selection problem. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pages 1362–1367.

[Lieder and Griffiths, 2020] Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1.

[Mercier and Sperber, 2017] Mercier, H. and Sperber, D. (2017). *The Enigma of Reason*. Harvard University Press.

[Morton, 2017] Morton, J. M. (2017). Reasoning under Scarcity. *Australasian Journal of Philosophy*, 95(3):543–559.

[Palmira, 2023] Palmira, M. (2023). Reasoning Simplifying Attitudes. *Episteme*, pages 1–14.

[Polonioli, 2014] Polonioli, A. (2014). Blame It on the Norm: The Challenge From ?Adaptive Rationality? *Philosophy of the Social Sciences*, 44(2):131–150.

[Rieskamp and Otto, 2006] Rieskamp, J. and Otto, P. E. (2006). SSL: A Theory of How People Learn to Select Strategies. *Journal of Experimental Psychology: General*, 135(2):207–236.

[Ross and Schroeder, 2014] Ross, J. and Schroeder, M. (2014). Belief, Credence, and Pragmatic Encroachment. *Philosophy and Phenomenological Research*, 88(2):259–288.

[Russell and Wefald, 1991] Russell, S. and Wefald, E. (1991). Principles of metareasoning. *Artificial Intelligence*, 49(1):361–395.

[Simon, 1978] Simon, H. A. (1978). Rationality as Process and as Product of Thought. *The American Economic Review*, 68(2):58–77.

[Simon, 2008] Simon, H. A. (2008). *The Sciences of the Artificial*. MIT Press, Cambridge, Mass., 3. ed., [nachdr.] edition.

[Staffel, 2013] Staffel, J. (2013). Can there be reasoning with degrees of belief? *Synthese*, 190(16):3535–3551.

[Staffel, 2019] Staffel, J. (2019). How Do Beliefs Simplify Reasoning? *Noûs*, 53(4):937–962.

[Thorstad, 2023] Thorstad, D. (2023). Why bounded rationality (in epistemology)? *Philosophy and Phenomenological Research*, page phpr.12978.

[Vul et al., 2014] Vul, E., Goodman, N., Griffiths, T. L., and Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science*, 38(4):599–637.