

Simplified Reasoning as Boundedly Rational

Sebastián Sánchez-Martínez ¹ | TU Dresden

Workshop in Theoretical Philosophy, Graz

23 May, 2025

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

Thanks for your feedback!

Plan:

1. Characterization
2. The Framework
3. Function of Simplified Reasoning
4. Normative Robustness
5. Wrap Up

GOAL: Spell out the normative consequences of approaching simplified reasoning from a tractability perspective.

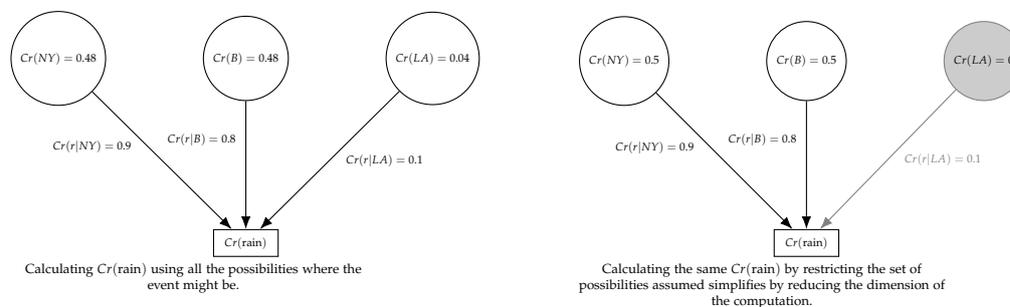
FRAMEWORK: Evaluate simplified reasoning using a simple yet powerful cost-benefit analysis. This is the consequentialist framework. Carefully justify: which policies of reasoning are available to limited agents, and what is the normative status of the verdicts.

UPSHOT: What bounded rationality demands from limited agents (the 'ought' of bounded rationality) requires substantial effort from limited agents without predicting widespread irrationality.

1. The Problem

Classical example (Staffel 2019): Forecasting rain during an event. Given uncertainty of the event's location, evidence-based credences are assigned to possible candidates. To simplify, a strategy is to screen-off the unlikely candidates, renormalize credences, and use the rain likelihoods only of the most likely ones to compute a weighted average formula.

Below, representation of the example:



How to characterize simplified reasoning for developing normative models? Try to do it in a very general way.

I will assume the following characterization of simplified reasoning:²

Highlight 1 (SR). *An agent S simplifies reasoning if and only if, and because, she is disposed to reason with restricted assumption sets in order to make reasoning problems tractable.*

Intuitively, tractability refers to the *feasibility under growing information* of a given reasoning method.³

² See: Harsanyi 1985, Bratman 1992, Lance 1995, Holton 2008, Wedgwood 2012, Ross and Schroeder 2014, Staffel 2019, Dinges 2021, Palmira 2023.

³ Tractability is formally defined in computer science and in artificial intelligence approaches to cognition. See Van Rooij et al. (2019).

If you impose certain constraints on selecting between two hobbies—coffee *versus* gardening—, it’s reasonably easy to select one—e.g., select the cheapest hobby that the majority of your best friends also like.

If you instead try to go over all parameters of each hobby— variety of flavors, potential long-term hedonic satisfaction, ethical considerations, health issues, probability of match with friends, etc.—, the decision space can grow so rapidly that even the best computers in the world would not be able to optimize this problem in a person’s lifetime.

Kinds of things that SR aims to include:

1. Dimension Reduction: Simplifying by using less variables to input in a given reasoning problem. This includes having thresholds for screening-off epistemic possibilities or criteria for excluding options.⁴
2. Problem Substitution: Simplifying by targetting a simpler problem.⁵
3. Attitude Switch: Simplifying by reasoning with different attitudes. Examples: reasoning with rounded credences instead of strict credences, reasoning with beliefs instead of credences, reasoning acceptance rather than belief, etc.

SOME REMARKS to narrow my target

- The dispositions to reason in certain ways called *policies* (shorthand π).⁶
- Policies are the choice variables in the framework. When S reasons adopting π_1 instead of π_2 , the framework assumes that it is at least in principle possible to choose between them.⁷
- The outputs of simplified reasoning can be doxastic (beliefs, credences) or conative (intentions, plans).
- The norms on the cognitive process of simplifying reasoning differ from the norms of their outputs. In my project, I tackle the rationality of the process.⁸

MY TARGET topic today is the rationality of SR:

Question: What is rationally required from an agent disposed to simplify her reasoning?

2. The Framework

What is a good framework for developing normative models of SR?

The consequentialist framework gives a way to approach different cases of simplified reasoning.

⁴ For the cases of epistemic possibilities see Staffel (2019), Ross and Schroeder (2014), and Clarke (2013). For cases of excluding options see Tversky (1969), Payne, Bettman, and Johnson (1993), and Simon (1956). My term *assumption sets* covers both the epistemic possibilities and the options that agents use in their reasoning.

⁵ See Kahneman (2013, Ch. 8) on the distinction between target and heuristic questions. See Tetlock and Gardner (2015) on ‘fermi-izing’. See also Pólya (2004) on the distinction between ‘proposed problem’ and ‘auxiliary problem’.

⁶ My approach to policies is as cognitive-behavioral dispositions to reason in a certain way. Roughly: cognitive functions that take a set of stimulus conditions (mental states, perceptual inputs, etc.) and maps them to the adoption of a certain way of reasoning.

⁷ “...to reason with restricted assumption sets (π_1) instead of full-information sets (π_2)...”

⁸ Some epistemologists think that knowledge or truth are norms of belief. On my view, this doesn’t necessarily commit them to taking those epistemic goods as normative standards for the cognitive dispositions that make reasoning tractable.

It says that a reasoning policy is assessed in a cost-benefit analysis that gives a net value to each policy.

The policy of reasoning that maximizes this value is minimally interpreted as the one that guides a rational assessment of the agent at a situation.⁹

Highlight 2 (Interpretation Keys). *The most important questions for interpreting the framework are*

[A] *How do we set the policies that enter the cost-benefit analysis?*

[B] *How should we, as theorists, interpret the π^* , the best policy at a situation?*

Answer to [A]: Pay close attention to tractability, the function of simplified reasoning! (§3)

If we ignore that tractability as a constraint, we risk including intractable policies in the set of choice variables.

Answer to [B]: Prima facie, what an agent ought to do at a reasoning situation is interpreted as a guide for rationally assessing the agent. (§4)

In what follows, I give more careful answers to each question.

3. The Function of Simplified Reasoning

Why is tractability so central for interpreting a normative model?

(Given my notion of policy) Good strategy to argue for the centrality of tractability: show a general connection between tractable reasoning policies and what bounded rationality requires.

ARGUMENT.

[P1] Limited agents can simplify their reasoning making a resourceful use of their cognitive limitations.

[P1.1] Limited agents can reason in demanding situations by solving problems that seem difficult at first glance.¹⁰

[P1.2] Limited reasoners cannot reason as if they were computationally unlimited.¹¹

[P2] If limited agents were not rational under their bounds, they could not exhibit the reasoning capacities we observe in common situations.

[C] Therefore, limited agents are rational under their bounds.

IN SHORT, if we don't constrain the set of optional policies in reasoning to the tractable ones, then a plausible explanation of dispositions to adopt reasoning-simplifying policies.

RESULT: What limited agents ought to do = adopt the policy $\pi_i \in \Pi_S$ that best strike the balance between its utility and costs = What bounded rationality requires.

⁹ Formally, a reasoning policy in a set of reasoning policies tractable for subject S , $\pi_i \in \Pi_S$, is evaluated according to its net value $V(\pi_i) = U(\pi_i) - C(\pi_i)$ that trades-off the utilities and costs of using that policy. The policy that the agent ought to implement is the one that maximizes the net value:

$$\pi^* = \arg \max_{\pi_i \in \Pi_S} V(\pi_i)$$

This is a simple yet powerful approach advocated in the resource-rationality tradition (Chater and Oaksford 1999; Icard 2023; Lieder and Griffiths 2020; Griffiths, Lieder, and Goodman 2015). See discussion in Rich et al. (2020).

Primarily, this is theoretically fruitful for investigating rationality—but it also can help the agent to get advice and guide her improvement. There is flexibility interpreting the kind of 'ought' at issue, more in §4.

¹⁰ E.g.: Finding a route, allocate personal budgets, playing games, all associated with tasks that are computationally intractable in general. See Macgregor and Ormerod (1996) for evidence of good human performance in the TPS.

¹¹ Alternative (Cherniak 1992, p. 78): "immediate synthetic a priori intuition" or a "transcendental ego, outside space, time, causality, and so on". But: these capacities doesn't seem to count as explanatory.

Why is this argument compelling? → It helps to distinguish between:

Fact 1: Tractable policies impose costs ≠ Fact 2: Some sets of tractable policies can lead to bad predicaments. Sometimes: irrational.

In some situations: one would prefer to have better options. No matter what one chooses, one would end up complaining. To fruitfully theorize about bounded rationality, we need to differentiate that every policy of reasoning comes with costs and that sometimes no all available policies result in negative values.

4. Normative Robustness

Why verdicts of a consequentialist treatment are normatively robust?

BECAUSE, the consequentialist framework does double work by answering two questions:

i. How limited agents ought to reason?

Answer: The consequentialist framework: Among a set of tractable policies for reasoning, limited agents should select the best policy (i.e. the policy that provides the best net trade-off between costs and benefits).

ii. What is rationally required from limited agents?

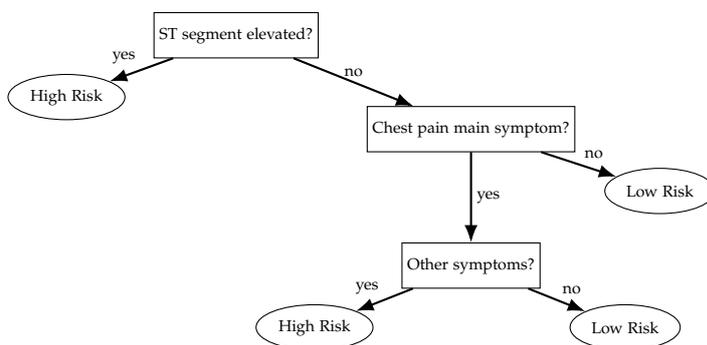
Answer: The consequentialist framework again!

THIS IS NOT WORDPLAY. Identifying what limited agents ought to do in reasoning with rational requirements is a key insight behind this strategy.

Interpretive constraint for π^* : it should capture genuine standards for normative assessment.¹²

How can the 'ought' of bounded rationality capture genuine norms? → Look at non-conventional standards of success. Example: predictive accuracy.

Example: Fast-and-frugal trees.



While complex models can be used effectively to fit a dataset, fast-and-frugal heuristics are robust in predicting new cases.

This is done by comparing how much the heuristic fits a known set of cases versus how many new cases it predicts. Each heuristic is compared to alternatives to determine which one better performs in prediction and fitting. Too much fitting indicates fitting of noise.

For a recent, similar view see Thorstad (2024).

For dispositionalist versions of reason-responsiveness see Wedgwood (2013) developing a proposal from Davidson (2001, Ch. 4).

Familiar from Lord (2017): what you ought to do and what you are rationally required to do are the same thing. Motivation: otherwise, it would be very difficult to answer to the question of why be rational? (Kolodny 2005)

¹² In the terms of Carr (2022), the 'ought' needs to be normatively robust (i.e. non-conventional and non-seriously context-sensitive).

Frugal = fewer cues used to make a categorization when compared, for instance, to logistic regression. Fast = fewer number of (logical or arithmetical) operations when compared to alternatives. See Martignon, Katsikopoulos, and Woike (2008).

The best reasoning policies are not just a convenient way (among alternatives) to coordinate actions, but they are adapted to predictability of cues. We use them as assessment principles.

What's important assessing a reasoning policy for their reliability is not whether the adoption of a reasoning policy succeeds or fails at a given case, but whether they succeed more than they fail across particular adoptions.

THE 'OUGHT' of bounded rationality captures features (like predictive accuracy) that matter in the kinds of situations for which certain reasoning policies are adapted.

In sum: simplified reasoning policies that are best at certain kinds of situations are often reliable. This reliability captures why they are normatively robust (why they capture genuine norms). This gives you different options to interpret the verdicts of a model, but they are constrained by ecological considerations: domains where policies normally work.

5. Wrap Up

Wrapping up...

Where does all of that leave us?

We have a general strategy for assessing agents who simplify their reasoning.

But we need to constrain how we set it up *and* how we interpret the verdicts.

There is a promising way of reconciling the "ought" of bounded rationality and demanding things that even limited agents are (boundedly) rationally required to do.

See Kornblith 1995, Ch. 5 and Kornblith 2021, Ch3.

What limited agents ought to do *does* provide a deontic ordering. It is just not global: it can be restricted to problem domains.

References

- Bratman, Michael E. (1992). "Practical Reasoning and Acceptance in a Context". In: *Mind* 101.401, pp. 1–16. DOI: [10.1093/mind/101.401.1](https://doi.org/10.1093/mind/101.401.1).
- Carr, Jennifer Rose (Dec. 2022). "Why Ideal Epistemology?" In: *Mind* 131.524, pp. 1131–1162. ISSN: 0026-4423, 1460-2113. DOI: [10.1093/mind/fzab023](https://doi.org/10.1093/mind/fzab023). (Visited on 05/07/2023).
- Chater, Nick and Mike Oaksford (Feb. 1999). "Ten Years of the Rational Analysis of Cognition". In: *Trends in Cognitive Sciences* 3.2, pp. 57–65. ISSN: 13646613. DOI: [10.1016/S1364-6613\(98\)01273-X](https://doi.org/10.1016/S1364-6613(98)01273-X). (Visited on 10/26/2023).
- Cherniak, Christopher (1992). *Minimal Rationality*. 2.print. A Bradford Book. Cambridge, Mass.: MIT Press. ISBN: 978-0-262-53087-3 978-0-262-03122-6.
- Clarke, Roger (2013). "Belief Is Credence One (in Context)". In: *Philosophers' Imprint* 13, pp. 1–18.
- Davidson, Donald (Sept. 2001). *Essays on Actions and Events*. 1st ed. Oxford University Press Oxford. ISBN: 978-0-19-924627-4 978-0-19-171519-8. DOI: [10.1093/0199246270.001.0001](https://doi.org/10.1093/0199246270.001.0001). (Visited on 05/19/2025).
- Dinges, Alexander (2021). "Beliefs Don'T Simplify Our Reasoning, Credences Do". In: *Analysis* 81.2, pp. 199–207. DOI: [10.1093/analysis/anaa084](https://doi.org/10.1093/analysis/anaa084).
- Griffiths, Thomas L., Falk Lieder, and Noah D. Goodman (2015). "Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic". In: *Topics in Cognitive Science* 7.2, pp. 217–229. ISSN: 1756-8765. DOI: [10.1111/tops.12142](https://doi.org/10.1111/tops.12142). (Visited on 08/17/2023).
- Harsanyi, John C. (1985). "Acceptance of Empirical Statements: A Bayesian Theory Without Cognitive Utilities". In: *Theory and Decision* 18.1, pp. 1–30. DOI: [10.1007/bf00134075](https://doi.org/10.1007/bf00134075).

- Holton, Richard (2008). "Partial Belief, Partial Intention". In: *Mind* 117.465, pp. 27–58. DOI: [10.1093/mind/fzn002](https://doi.org/10.1093/mind/fzn002).
- Icard, Thomas (2023). "Resource Rationality". (Visited on 02/01/2024).
- Kahneman, Daniel (2013). *Thinking, Fast and Slow*. 1st pbk. ed. New York: Farrar, Straus and Giroux. 499 pp. ISBN: 978-0-374-53355-7.
- Kolodny, Niko (July 2005). "Why Be Rational?" In: *Mind* 114.455, pp. 509–563. ISSN: 0026-4423. DOI: [10.1093/mind/fzi509](https://doi.org/10.1093/mind/fzi509). (Visited on 05/09/2025).
- Kornblith, Hilary (1995). *Inductive Inference and Its Natural Ground: An Essay in Naturalistic Epistemology*. 1. paperback ed. A Bradford Book. Cambridge, Mass.: MIT Press. ISBN: 978-0-262-11175-1 978-0-262-61116-9.
- (Dec. 2021). *Scientific Epistemology: An Introduction*. 1st ed. Oxford University Press New York. ISBN: 978-0-19-760955-2 978-0-19-760959-0. DOI: [10.1093/oso/9780197609552.001.0001](https://doi.org/10.1093/oso/9780197609552.001.0001). (Visited on 03/24/2025).
- Lance, Mark Norris (1995). "Subjective Probability and Acceptance". In: *Philosophical Studies* 77.1, pp. 147–179. DOI: [10.1007/bf00996316](https://doi.org/10.1007/bf00996316).
- Lieder, Falk and Thomas L. Griffiths (2020). "Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources". In: *Behavioral and Brain Sciences* 43, e1. ISSN: 0140-525X, 1469-1825. DOI: [10.1017/S0140525X1900061X](https://doi.org/10.1017/S0140525X1900061X). (Visited on 07/03/2023).
- Lord, Errol (2017). "What You?Re Rationally Required to Do and What You Ought to Do". In: *Mind* 126.504, pp. 1109–1154. DOI: [10.1093/mind/fzw023](https://doi.org/10.1093/mind/fzw023).
- Macgregor, J. N. and T. Ormerod (June 1996). "Human Performance on the Traveling Salesman Problem". In: *Perception & Psychophysics* 58.4, pp. 527–539. ISSN: 0031-5117, 1532-5962. DOI: [10.3758/BF03213088](https://doi.org/10.3758/BF03213088). (Visited on 05/19/2025).
- Martignon, Laura, Konstantinos V. Katsikopoulos, and Jan K. Woike (Dec. 2008). "Categorization with Limited Resources: A Family of Simple Heuristics". In: *Journal of Mathematical Psychology* 52.6, pp. 352–361. ISSN: 00222496. DOI: [10.1016/j.jmp.2008.04.003](https://doi.org/10.1016/j.jmp.2008.04.003). (Visited on 10/12/2023).
- Palmira, Michele (2023). "Reasoning Simplifying Attitudes". In: *Episteme*, pp. 1–14. DOI: [10.1017/epi.2023.4](https://doi.org/10.1017/epi.2023.4).
- Payne, John W., James R. Bettman, and Eric J. Johnson (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press. ISBN: 978-0-521-41505-7. DOI: [10.1017/CB09781139173933](https://doi.org/10.1017/CB09781139173933). URL: <https://www.cambridge.org/core/books/adaptive-decision-maker/C2F0579B685EC397059F5D386E7B2045> (visited on 04/05/2023).
- Pólya, George (2004). *How to Solve It: A New Aspect of Mathematical Method*. Expanded Princeton Science Library ed. Princeton Science Library. Princeton [N.J.]: Princeton University Press. ISBN: 978-0-691-11966-3.
- Rich, Patricia et al. (2020). "How Intractability Spans the Cognitive and Evolutionary Levels of Explanation". In: *Topics in Cognitive Science* 12.4, pp. 1382–1402. ISSN: 1756-8765. DOI: [10.1111/tops.12506](https://doi.org/10.1111/tops.12506). (Visited on 10/04/2023).
- Ross, Jacob and Mark Schroeder (2014). "Belief, Credence, and Pragmatic Encroachment". In: *Philosophy and Phenomenological Research* 88.2, pp. 259–288. DOI: [10.1111/j.1933-1592.2011.00552.x](https://doi.org/10.1111/j.1933-1592.2011.00552.x).
- Simon, H. A. (1956). "Rational Choice and the Structure of the Environment". In: *Psychological Review* 63.2, pp. 129–138. ISSN: 1939-1471. DOI: [10.1037/h0042769](https://doi.org/10.1037/h0042769).
- Staffel, Julia (2019). "How Do Beliefs Simplify Reasoning?" In: *Noûs* 53.4, pp. 937–962. DOI: [10.1111/nous.12254](https://doi.org/10.1111/nous.12254).
- Tetlock, Philip E and Dan Gardner (2015). *Superforecasting: The Art and Science of Prediction*. Crown Publishing.
- Thorstad, David (June 2024). *Inquiry Under Bounds*. 1st ed. Oxford University Press Oxford. ISBN: 978-0-19-888614-3 978-0-19-199422-7. DOI: [10.1093/oso/9780198886143.001.0001](https://doi.org/10.1093/oso/9780198886143.001.0001). (Visited on 08/22/2024).
- Tversky, Amos (1969). "Intransitivity of Preferences". In: *Psychological Review* 76.1, pp. 31–48. ISSN: 1939-1471. DOI: [10.1037/h0026750](https://doi.org/10.1037/h0026750).

- Van Rooij, Iris et al. (2019). *Cognition and Intractability: A Guide to Classical and Parameterized Complexity Analysis*. Cambridge: Cambridge University Press. ISBN: 978-1-107-04399-2.
- Wedgwood, Ralph (2012). "Outright Belief". In: *Dialectica* 66.3, pp. 309–329. DOI: [10.1111/j.1746-8361.2012.01305.x](https://doi.org/10.1111/j.1746-8361.2012.01305.x).
- (2013). "Rational 'Ought' Implies 'Can'". In: *Philosophical Issues* 23.1, pp. 70–92. ISSN: 1758-2237. DOI: [10.1111/phis.12004](https://doi.org/10.1111/phis.12004). (Visited on 05/02/2025).