UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
How hard is cognitive science?

Permalink
https://escholarship.org/uc/item/8cr8x1c4

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Rich, Patricia
de Haan, Ronald
Wareham, Todd
et al.

Publication Date
2021

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

How hard is cognitive science?

Patricia Rich (patricia.rich@uni-bayreuth.de)
Philosophy Department, University of Bayreuth, Germany

Ronald de Haan (me@ronalddehaan.eu)
Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands

Todd Wareham (harold@mun.ca)
Department of Computer Science, Memorial University of Newfoundland, Canada

Iris van Rooij (i.vanrooij@donders.ru.nl)
Donders Institute for Brain, Cognition, and Behaviour, Radboud University, The Netherlands

Abstract

Cognitive science is itself a cognitive activity. Yet, computational cognitive science tools are seldom used to study (limits of) cognitive scientists' thinking. Here, we do so using computational-level modeling and complexity analysis. We present an idealized formal model of a core inference problem faced by cognitive scientists: Given observations of a system's behaviors, infer cognitive processes that could plausibly produce the behavior. We consider variants of this problem at different levels of explanation and prove that at each level, the inference problem is intractable, or even uncomputable. We discuss the implications for cognitive science.

Keywords: philosophy of cognitive science, formal epistemology, abduction, computational complexity, intractability

Introduction

Imagine a scientist, called Dr. Conjectura, who wants to explain some cognitive system's behavior and inner workings. They may be interested in explanations at different levels of granularity, ranging from high-level functional explanations of its input-output behavior to lower-level process or mechanistic explanations (Marr, 1982; Bechtel & Shagrir, 2015; Anderson, 1990; Egan, 2017; Jarecki, Tan, & Jenny, 2020; Love, 2020). A number of challenges facing Dr. Conjectura have been recognized and studied to various degrees. Traditionally, challenges posed by uncertainty have been the focus (e.g., statistically inferring effects from data, problems of induction and generalizability, underdetermination of theory by data). In this paper, we show that even if all of this uncertainty were removed, a major obstacle to explaining cognition would remain. In particular, the inference problems that Dr. Conjectura aims to solve are computationally intractable, and some are even uncomputable. This shows that we cannot hope to explain cognition solely through techniques for removing or responding to uncertainty; the task is difficult in a more fundamental way, and different kinds of strategies are necessary to address this.

Based on where the most energy is expended, one would think that the main challenges for cognitive science have to do with gathering sufficient, high-quality empirical evidence. This idea is further strengthened by the fact that recent science reforms (motivated in large part by the 'replication crisis') have focused on devising practices and procedures geared towards reducing uncertainty about which effects are real, and which ones illusory (Open Science Collaboration, 2015; Nosek et al., 2019). Whether or not these approaches help or hinder is a topic of debate (Szollosi et al., 2020; Devezer, Navarro, Vandekerckhove, & Ozge Buzbas, 2021; Irvine, 2021), but let's assume for sake of argument that this challenge were resolved, and our experiments and statistical analyses would yield only true results. We would still have a major problem, namely the problem of induction (Hume, 1739; Goodman, 1983), since present results are insufficient to infer future observations. Dr. Conjectura may have observed people behaving a certain way or a particular brain region being activated during an experiment, but remain uncertain about how to generalize these observations across contexts and time.

But let's suppose that the problem of induction were also solved; let's even imagine that Dr. Conjectura has full, automatic access to all facts, present and future. Then their predictions would be accurate, a fortiori, but would Dr. Conjectura also be able to supply us with correct explanations of cognitive phenomena? It might seem that little work would remain, but explanation might still prove elusive. Not only would the sheer volume of data present a practical challenge, but theory is (notoriously) underdetermined by data (Quine, 1951).[1] Dr. Conjectura may provide an explanation only for future researchers to replace it with an alternative that they find preferable not because it is objectively more accurate, but because it is e.g. simpler or more fruitful (Kuhn, 1962).

Now, let's imagine that even this challenge would disappear; we somehow knew for sure that only one acceptable theory were consistent with our data. We suspect that many would think that no serious theoretical challenges remained, and that Dr. Conjectura could certainly provide all of the explanations we could want. The purpose of this paper is to demonstrate that this is not true: even if all uncertainty is removed from scientific inference problems, there are further principled barriers to deriving explanations, resulting from the computational complexity of the inference problems.This implies that our inferential challenges cannot be solved purely by improving our responses to uncertainty. Furthermore, it would be a mistake to think that finding strategies to address

[1] The *identifiability problem* in cognitive science (Anderson, 1991; Varma, 2014) is a special case.

computational obstacles to scientific inference can be postponed until after the other challenges have been addressed; methodological proposals must be assessed in terms of their prospects all-things-considered, lest Dr. Conjectura waste time on short-sighted measures that cannot lead to long-term success.

The remainder of this paper is organized as follows. We first present idealized models of the inference problems Dr. Conjectura faces when generating explanations of the workings of a cognitive system. Next, we analyze how hard these inference problems are using concepts, tools and techniques from the mathematical theories of computability and complexity. Finally, we position our conclusions with respect to existing work and use a fictional dialogue to explain the implications of our results for practicing cognitive scientists.

Formalizing scientific abductive inference

In this section, we develop formalizations of inference problems posed to Dr. Conjectura when they try to explain the workings of a cognitive system M (see Fig. 1). Without loss of generality[2], we work with a highly idealized scenario: Dr. Conjectura has access to a data set $D \subseteq S \times B$ where all observations are free from error in measurement or interpretation. Each pair $(s,b) \in D$ denotes an observed behavior $b \in B$ of M in a situation $s \in S$ (see Table 1).

Table 1: Illustration of a hypothetical data set D. Here, situations can be anything (e.g., choice problems, visual tasks, social situations, or combinations thereof), and behavior can also be anything (e.g., choices, movement trajectories, speech acts, reaction times, or combinations thereof).

entry nr.	situation	behavior
1	s_1	b_3
2	s_1	b_4
3	s_2	b_3
4	s_6	b_1
5	s_7	b_1
...	...	...
...	...	...
...	...	...
365	s_{63}	b_3

Dr. Conjectura aims to generate explanations of M from D (a.k.a. abductive inference) at a functional level and/or algorithmic level while taking into account background knowledge and assumptions about the nature of the world (e.g., that computation takes time, that physical systems are bounded in space) and the system under study (e.g., M belongs to certain class of mechanisms). We formalize this idea as follows: Dr. Conjectura seeks as a functional level explanation a function $F \in \mathcal{F}$, where $\mathcal{F}$ is a class of functions $F : S \to 2^B$ that

[2]See the Discussion section.

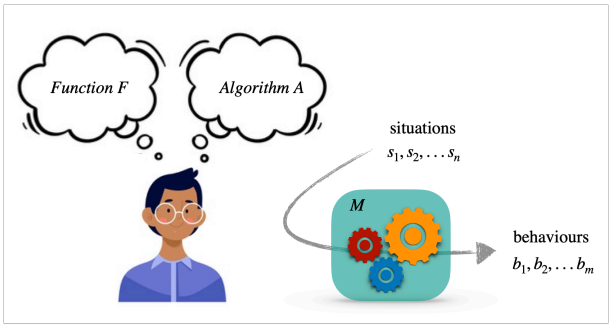

Figure 1: A (cognitive) system M computes an unknown function F_M using an unknown algorithm A_M. Dr. Conjectura observes the behavior of M in various situations and tries to come up with a function F and an algorithm A that are consistent with the observations and background knowledge. [Figure created with elements from freepik.com]

satisfy these background assumptions. Similarly, Dr. Conjectura seeks as an algorithmic level explanation an algorithm $A \in \mathcal{A}$, where $\mathcal{A}$ is a class of algorithms able to compute functions $F \in \mathcal{F}$ in a way that satisfies the relevant background assumptions.

Dr. Conjectura will need to be able to describe these explanations F and A somehow, possibly using a mix of natural language, mathematical notation and diagrammatic sketches (Guest & Martin, 2021; van Rooij & Blokpoel, 2020). We formalize the scientific language system that Dr. Conjectura uses for functional level explanations as a function $\mathcal{L}_\mathcal{F}$ that maps any string L_F (a description of a function), and sets S and B, to a function $\mathcal{L}_\mathcal{F}(S,B,L_F) = F : S \to 2^B$. In practice, such desciptions cannot be arbitrarily long, but need to fit a scientific article or book of reasonable length. Therefore we assume some upper bound K on the length of descriptions that Dr. Conjectura will consider workable and publishable.

With these formalizations in hand we can now more precisely define the inference problem that Dr. Conjectura solves as they come up with functional level explanations (we consider an algorithmic level variant later on):

$(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE
Given: A data set D with observed situation-behavior pairs $(s,b) \in S \times B$, generated by an unknown function $F_M : S_M \to 2^{B_M}$ of type $\mathcal{F}$, where $S \subseteq S_M$, $B \subseteq B_M$, and an upper bound $K \in \mathbb{N} \cup \{\infty\}$.
Inference: A description $L_F \in \mathcal{L}_\mathcal{F}$ of a function $F \in \mathcal{F}$ that is *consistent* with D, such that L_F has length at most K, if such an L_F exists. "None" otherwise.

What is meant here by 'consistent' is somewhat open to interpretation, and different interpretations yield slightly different variants of the inference problem, just as different classes $\mathcal{F}$ and languages $\mathcal{L}_\mathcal{F}$ do. For ease of presentation, let's take 'consistent' to mean that for each $(s,b) \in D$, $b \in F(s)$. As we

explain in the Discussion, our analyses and results are robust to variations in the definition.

We illustrate the abductive inference problem with an example: Say Dr. Conjectura is interested in explaining human choice behavior. A colleague, Dr. Mensura, has provided a large data set $D \subseteq \mathbb{S} \times \mathbb{B}$, where the $s \in \mathbb{S}$ are situations in which individuals are presented with choices (i.e., sets of options that the person can choose from) and each $b \in \mathbb{B}$ is a particular choice made by the person (one or more of the options from the given set). Based on their previous training in behavioral economics, Dr. Conjectura postulates that the functional level explanation L_F describing participant behavior is to be built with the following constraints ($\mathcal{F}$): each person p has a utility function $u_p : X \to \mathbb{R}$ that maps choice options $x \in X$ to subjective values $u_p(x) \in \mathbb{R}$, and a person selects options that meet some minimum subjective value criterion t. I.e., given a subset $X' \subseteq X$ of options, person p will choose an option $x \in \{x | x \in X' \text{ and } u_p(x) \geq t\}$. Given this background commitment, the abductive inference problem for a given D comes down to searching for a combination of functions u_p and values t_p that is consistent with D. Of course, this explanation would also need to be described somehow (in $\mathcal{L}_\mathcal{F}$, e.g. with a mix of natural language and formal notation, as we ourselves used above), and this description should not be too long to be practical.

This is but one example of assumptions about $\mathcal{F}$ that Dr. Conjectura could make. They could make additional assumptions (e.g., that the utility functions satisfy rationality principles like transitivity); very different assumptions (e.g., that preferences are dynamically constructed and not describable by any option-level utility function (Payne, Bettman, & Johnson, 1993)); or almost no assumptions (e.g., that decision making is describable by some (computable or tractable) function (van Rooij, Blokpoel, Kwisthout, & Wareham, 2019)). This illustrates that $(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE can be interpreted very narrowly as model fitting, or very broadly as model specification or theory formation.[3]

We next define the algorithmic-level version of Dr. Conjectura's abductive inference problem. Again, they will need a language ($\mathcal{L}_\mathcal{A}$) to express their explanations, which can be of any level of granularity, ranging from high (e.g. cognitive) to low (neural). The main difference with the functional-level language ($\mathcal{L}_\mathcal{F}$) is that $\mathcal{L}_\mathcal{A}$ must be *constructive*. For instance, $\mathcal{L}_\mathcal{F}$ allowed Dr. Conjectura to postulate that a decision maker presented with a set of options X' chooses any $x \in \{x | x \in X'$ and $u_p(x) \geq t\}$, without specifying *how* this x is found. In contrast, an algorithmic level explanation needs to specify exactly how a person is believed to *compute* set membership. $\mathcal{L}_\mathcal{A}$ can be thought of as akin to a programming language, but allowing only those computational steps that Dr. Conjectura believes are realizable by the system under study. Like functional level explanations, algorithmic level explanations must be publishable, and hence cannot be arbitrarily large (although the domains over which they compute can be infinite). We can thus define Dr. Conjectura's problem of inferring algorithmic level explanations as follows:

$(\mathcal{A}, \mathcal{L}_\mathcal{A})$-ABDUCTIVE INFERENCE
Given: A data set D with observed situation-behavior pairs $(s,b) \in S \times B$, generated by an unknown algorithm A_M of type $\mathcal{A}$ computing an unknown function $F_M : S_M \to 2^{B_M}$, where $S \subseteq S_M$, $B \subseteq B_M$, and an upper bound $K \in \mathbb{N} \cup \{\infty\}$.
Inference: A description $L_A \in \mathcal{L}_\mathcal{A}$ of an algorithm $A \in \mathcal{A}$ that computes a function F_A that is *consistent* with D, such that L_A has length at most K, if such an L_A exists. "None" otherwise.

Note that $(\mathcal{A}, \mathcal{L}_\mathcal{A})$-ABDUCTIVE INFERENCE and $(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE really are two distinct problems; a solution to one does not automatically yield a solution to the other. In one direction this may be obvious: a description L_F of a function F need not specify an algorithm A for computing F. To see that the reverse also holds, note that a description L_A of an algorithm A may concisely specify a method for computing a function F, but that in itself does not naturally give a short description of F in the explanatory language $\mathcal{L}_\mathcal{F}$ (Bechtel & Shagrir, 2015; Egan, 2017; Varma, 2014).

Computational complexity analysis

To analyse how hard the inference problems posed to Dr. Conjectura are, we build on concepts and proof techniques from the mathematics of computability and complexity. We start by presenting key definitions in an accessible form that suffices for our purposes. For more extensive and formal treatments see textbooks such as (Lewis & Papadimitriou, 1997; Garey & Johnson, 1979; Arora & Barak, 2009; Primiero, 2019; van Rooij et al., 2019).

Definition 1 (Computability) A relation (e.g., function or inference problem) $Q : X \to 2^Y$ is said to be *computable* if there exists at least one algorithm that can compute $y \in Q(x)$ for any $x \in X$. Otherwise we say that Q is *uncomputable*.

Definition 2 (P-time algorithm) An algorithm A is said to be a *polynomial-time* algorithm if it runs in time $O(n^c)$, where n is a measure of the input size and c is a constant.

Algorithms that take more than polynomial time, such as exponential time algorithms, are generally regarded as *intractable* for all but small input sizes (Garey & Johnson, 1979; Arora & Barak, 2009). To illustrate why, consider an abductive inference problem with a data set D, of modest size $|D| = 100$. An exponential time algorithm running in, say, time $O(c^n)$ with $c = 2$ would take on the order of 10^{30} steps, which is more than the number of seconds that passed since the birth of the universe ($< 10^{18}$ seconds). For $|D| = 500$, this

[3] Dr. Conjectura can also drop assumptions about $\mathcal{F}$ whenever a previous inference returned "none," until inference is possible. We do not model the problem of deciding which assumptions to drop. Hence, our hardness results are really *lower*-bounds on complexity.

number would be on the order of 10^{150}, which far surpasses the number of atoms in the universe ($< 10^{82}$ atoms). This means that for such a data set, even a brain or machine—or a collection of brains or machines, with as many parallel computing channels as there are atoms in the universe—may take as long as the time that has passed since the birth of the universe to complete the inferential process. Suffice it to say that in practice, intractable abductive inferences are not feasible for medium to large inputs.

Definition 3 (Tractability) A relation Q is said to be *tractable* if there exists at least one polynomial-time algorithm that can compute it. Otherwise we say Q is *intractable*.

Definition 4 (Tractable verifiability) A relation $Q : X \rightarrow 2^Y$ is said to be *tractably verifiable* if there exists an algorithm that (i) runs in polynomial time in the size of x, and (ii) given x and y, can verify that $y \in Q(x)$ holds.

Note that if Q is tractable then it is also tractably verifiable. The converse is generally assumed to be false (Goldreich, 2010, Chapter 2).

Here we will assume that both the functions computed by cognitive systems and the inferential capacities of cognitive scientists themselves are bounded by the requirement of tractability (for an extensive argument see (van Rooij et al., 2019)). Given this assumption, we ask: How hard is cognitive science? To address this question, we derive a set of theorems that we collectively refer to as *Conjectura theorems* (see the supplementary materials[4] for details and proofs and Table 2 for an overview):

Theorem 1 If there is no bound on the length of the explanation $L_F \in \mathcal{L}_\mathcal{F}$ (i.e., $K = \infty$), then $(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE is uncomputable for some $\mathcal{L}_\mathcal{F}$.

Theorem 1 holds even if $\mathcal{L}_\mathcal{F}$ is tractable (and hence $\mathcal{F}$ contains only tractable functions). This shows that the uncomputability of $(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE is not due to excessive complexity of the to-be-explained cognitive system, but rather to the absence of the bound on the length of explanations. This interpretation is confirmed by Theorem 2.

Theorem 2 If there is a bound $K \in \mathbb{N}$ on the length of the explanation $L_F \in \mathcal{L}_\mathcal{F}$, then $(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE is computable for all $\mathcal{L}_\mathcal{F}$.

While a bound on the length of explanations buys computability, it does not yet buy tractability, as shown next.

Theorem 3 If there is a bound $K \in \mathbb{N}$ on the length of the explanation $L_F \in \mathcal{L}_\mathcal{F}$ and $\mathcal{L}_\mathcal{F}$ is tractable (and hence $\mathcal{F}$ contains only tractable functions), then $(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE is intractable for some $\mathcal{L}_\mathcal{F}$.

[4] Available on OSF at https://osf.io/gpkhj/

Theorem 3 shows that there cannot exist any polynomial-time algorithm for generating tractable functional explanations (of bounded size) for any given data set. What is possible, however, is to tractably recognize explanations when stumbled upon by chance, as shown next.

Theorem 4 If there is a bound $K \in \mathbb{N}$ on the length of the explanation $L_F \in \mathcal{L}_\mathcal{F}$ and $\mathcal{L}_\mathcal{F}$ is tractable (hence $\mathcal{F}$ contains only tractable functions), then $(\mathcal{F}, \mathcal{L}_\mathcal{F})$-ABDUCTIVE INFERENCE is tractably verifiable for all $\mathcal{L}_\mathcal{F}$.

Theorem 4 holds even if M is assumed to be a limited computational device, such as a finite state automaton.

The theorems listed so far pertain to the abduction of functional level explanations. We have a fully analogous set of theorems for algorithmic level abduction.

Theorem 5 If there is no bound on the length of the explanation $L_A \in \mathcal{L}_\mathcal{A}$ (i.e., $K = \infty$), then $(\mathcal{A}, \mathcal{L}_\mathcal{A})$-ABDUCTIVE INFERENCE is uncomputable for some $\mathcal{L}_\mathcal{A}$.

Theorem 6 If there is a bound $K \in \mathbb{N}$ on the length of the explanation $L_A \in \mathcal{L}_\mathcal{A}$, then $(\mathcal{A}, \mathcal{L}_\mathcal{A})$-ABDUCTIVE INFERENCE is computable for all $\mathcal{L}_\mathcal{A}$.

Theorem 7 If there is a bound $K \in \mathbb{N}$ on the length of the explanation $L_A \in \mathcal{L}_\mathcal{A}$ and $\mathcal{L}_\mathcal{A}$ is tractable (and hence $\mathcal{A}$ contains only tractable functions), then $(\mathcal{A}, \mathcal{L}_\mathcal{A})$-ABDUCTIVE INFERENCE is intractable for some $\mathcal{L}_\mathcal{A}$.

Theorem 8 If there is a bound $K \in \mathbb{N}$ on the length of the explanation $L_A \in \mathcal{L}_\mathcal{A}$, $\mathcal{L}_\mathcal{A}$ is tractable (hence $\mathcal{A}$ contains only tractable functions), then $(\mathcal{A}, \mathcal{L}_\mathcal{A})$-ABDUCTIVE INFERENCE is tractably verifiable for all $\mathcal{L}_\mathcal{A}$.

Discussion

We started our exploration by asking how hard cognitive science would be if there were no uncertainty in our observations and we had all the relevant data. The Conjectura theorems show that even in such an ideal situation, generating explanations consistent with our observations defies any efficient abductive inference procedure (see Table 2 for an overview of the results). We situate our results in the literature before drawing out their implications.

While the results will strike some of the cognitive science community as unintuitive and disappointing, others will not be surprised because there is a family of negative theoretical results which are similar in spirit, coming from multiple research traditions and dating back to at least the 1960's. Most salient are computational and formal learning theory (Solomonoff, 1964; Angluin, 1992; Kelly, 1996; Gierasimczuk, 2010) and the cognitive science of science (Thagard, 2012). The lines of research differ in their aims, degree of generality, and the types of scientific problems they address. There are several key differences between our formalization and those existing approaches. We focus on explanation

Table 2: How hard is cognitive science?

Type of explanation	Assumptions	Computable	Tractable	Tractably verifiable
Functional	$\mathcal{L}_{\mathcal{F}}$ tractable, $K = \infty$	No	No	No
Algorithmic	$\mathcal{L}_{\mathcal{A}}$ tractable, $K = \infty$	No	No	No
Functional	$\mathcal{L}_{\mathcal{F}}$ tractable, $K \in \mathbb{N}$	Yes	No	Yes
Algorithmic	$\mathcal{L}_{\mathcal{A}}$ tractable, $K \in \mathbb{N}$	Yes	No	Yes

instead of, for instance, induction, prediction or testability (Kelly, 1996; Kelly & Schulte, 1995). We investigate computational complexity rather than learnability in the limit (Gold, 1967; Solomonoff, 1964). We model the specific problem of abduction given background assumptions about the cognitive system, in contrast to more generic formalisms (Gold, 1978; Angluin, 1992). Our models are, however, general enough to encompass all explanatory computational frameworks in cognitive science, in contrast to Bayesian (Kwisthout, 2011; Lipton, 2003) or coherentist (Thagard, 1989; Thagard & Verbeurgt, 1998) models of abduction. In sum, our formalisms allow us to show how hard specific, recognizable abductive inference problems of interest to cognitive scientists are.

It is important to show which negative results apply to these specific problems, even though the nature of negative results is that there can be no "solution" in the form of a way to get around them. Compare to the "no free lunch theorem" of machine learning (Wolpert, 1996), which says essentially that there exists no best algorithm for solving all learning problems; awareness of this is important so that researchers don't go looking for a generic best algorithm. Similarly, it is important to generate awareness among all cognitive scientists that there can be no fixed, general procedure to generate good explanations, precisely so that such a procedure is not sought, and claims of its existence are not trusted.

Unpacking the full implications of our results for cognitive science research practice is not easy, and we imagine that those reading this may have all kinds of questions, objections, or counter-intuitions. Given space limitations, we cannot possibly address them all. Instead, we unpack the implications of the results using a fictive dialogue, addressing the most likely concerns along the way. In the dialogue, Dr. Conjectura (denoted by **C**) plays the role of the skeptic who does not see the relevance of the results for their own practices. **R** relays our responses.

C: I appreciate you trying to help me achieve my research goals, but I can't see how you are doing so. How are the theorems relevant to me? I am never in that ideal situation.

R: What ideal situation?

C: You formalized my inferential problems by assuming I have perfect, errorless observations. But my data are always incomplete and noisy.

R: The theorems show that in the ideal situation finding explanations consistent with the facts is not tractable. How can more uncertainty about the relevant facts make this problem *easier*? It seems it can only make it *harder*.

C: Fair. But you set an unrealistic standard for explanation. No explanation is perfect, but at best an approximation.

R: What do you mean exactly by 'approximation'?

C: Well, for instance, explanations do not always need to be consistent with *all* the data.

R: We need not assume such a high standard. Even if an explanation needs to be consistent with, say, half of the data,[5] generating such 'half-consistent' explanations remains intractable.

C: Oh. That's counter-intuitive.

R: I hope this takes away your worries about the idealizations we introduced? In general, many problems that are intractable to solve exactly are also hard to solve approximately, for various meanings of 'approximation.'[6]

C: But I still do not understand. If you would just give me perfect, error-free observations, shouldn't it be easy for me to infer the mechanism producing that data?

R: Explanation does not come for free. The number of possible mechanisms you *could* describe with language and mathematics is astronomical. Finding a description that pinpoints a mechanism consistent with the data is like finding a needle in a haystack: there exists no general efficient procedure for searching the space.

C: But I've already narrowed down the options. I'm looking only for explanations of a particular cognitive architecture type: [insert your favorite framework, e.g., ACT-R, Adaptive Toolbox, PDP, Subsumption-Architecture, etc.].

R: Our analyses encompass this view, as one option, by constraining the space of possible functions (the set $\mathcal{F}$) and algorithms ($\mathcal{A}$), according to your architectural commitments. Even with such general *a priori* commitments, the space remains astronomically large for architectures with non-trivial computing power.

C: What do you mean by non-trivial?

[5] Equivalently, one could assume that D contains only those data you care about and think are relevant for ones explanation.

[6] See e.g. (Arora, 1998; Garey & Johnson, 1979).

R: Well, even if a system has few possible internal states and its behavior is fully governed by simple rules, generating explanations of its behaviors remains intractable.[7] Do you think that human cognition is simpler than this?

C: No, likely more complex.

R: Then our intractability results apply to your work.

C: Are you saying my work is hopeless? I cannot hope to ever generate a satisfactory explanation for cognition?

R: I wouldn't say hopeless. If you were to hit upon a satisfactory explanation through sheer luck, then you could recognize this.[8]

C: Sigh. That's not much of a plan ...

R: I don't think you need to be any more discouraged by intractability than by the inherent uncertainty in your data, generalizations, and theory that you were already dealing with. But it does mean that your inferential work cannot be proceduralized in any efficient way. So best not try to make an algorithm, or an otherwise too-strict set of rules, to replace your scientific thinking.

C: Why not? What could go wrong?

R: You may fool yourself into thinking you are searching the whole space, while you are actually stuck in a small corner of an astronomical space outside your consideration. It may also cause you to assume that the system you are studying is simpler than you really believe, because otherwise your procedures would not converge efficiently.

C: Well if any procedure I might use will hold me back, what can I do?

R: I would endorse a meta-approach of not proceduralizing. This is especially important now, as we increasingly focus on a too narrow set of methodological approaches in cognitive science.[9] The best advice I can give pertains to the community: our only hope of understanding the mind is if the community allows for pluralism[10] in approaches and an unbounded number of procedures different researchers may adopt.

C: Why unbounded?

R: Because it is known that intractable problems cannot be solved by a *fixed* number of parallel procedures.[11]

C: But if we impose no limit on the number of approaches, wouldn't there be many bad ones?

R: Recognizing the need for and legitimacy of alternative approaches is a prerequisite to productive critique.[12] So you can critique approaches on substantive grounds, but I must dissuade you from viewing any fixed (set of) procedure(s) as the right one and trying to convince others that they should adopt it too. I've noticed you grumbling about the too-subjective methods[13] some of your colleagues are using, and I must encourage you to live and let live.[14]

Acknowledgments

The authors thank the anonymous reviewers for helpful comments on a previous version of this paper and Berna Devezer for discussions that have inspired this research. TW was supported by NSERC Discovery Grant 228104-2015. IvR acknowledges Schloss Dagstuhl: Leibniz Centre for Informatics for the support of a research retreat (nr. 19299) and the Netherlands Institute for Advanced Studies in the Humanities and Social Sciences (NIAS-KNAW) and the Lorentz Center for a 2020/21 Distinguished Lorentz Fellowship.

References

Anderson, J. (1990). *The adaptive character of thought*. Psychology Press.

Anderson, J. (1991). Is human cognition adaptive? *Behavioral and brain sciences*, *14*(3), 471–571.

Angluin, D. (1992). Computational learning theory: survey and selected bibliography. In *Proceedings of 24th ACM symposium on Theory of Computing* (pp. 351–369).

Arora, S. (1998). The approximability of NP-hard problems. In *Proceedings of the thirtieth annual acm symposium on theory of computing* (pp. 337–348).

Arora, S., & Barak, B. (2009). *Computational complexity: a modern approach*. Cambridge University Press.

Bechtel, W., & Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information-processing mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322.

Dale, R. (2008). The possibility of a pluralist cognitive science. *Journal of Experimental and Theoretical Artificial Intelligence*, *20*(3), 155–179.

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PLOS ONE*, *14*(5), e0216125.

Devezer, B., Navarro, D. J., Vandekerckhove, J., & Ozge Buzbas, E. (2021). The case for formal methodology in scientific reform. *Royal Society Open Science*, *8*(3), 200805.

[7] See Thm A.1 and Proposition A.3 in Supplementary materials.

[8] By Thms. 3, 6. Note that so far, we only know this for ideal situations as per our problem definitions.

[9] For instance, due to increasing dominance of experimental psychology in cognitive science over the last 30 years (Gentner, 2019).

[10] Cf. Devezer, Nardin, Baumgaertner, and Buzbas (2019) and Dale (2008).

[11] By Lemma 3, p. 481 of van Rooij et al. (2012).

[12] Cf. Dow (2008).

[13] Cf. Field and Derksen (2021).

[14] Cf. Feyerabend (1975). A similar moral can be drawn from the recent literature in the social epistemology of science, which emphasizes the group-level benefits of various kinds of scientific diversity; see e.g. (Kitcher, 1990; Weisberg & Muldoon, 2009; Thoma, 2015; Zollman, 2010; O'Connor & Bruner, 2019).

Dow, S. C. (2008). Plurality in orthodox and heterodox economics. *Journal of Phil. Economics*, *1*(2), 73–96.

Egan, F. (2017). Function-theoretic explanation and the search for neural mechanisms. In *Explanation and Integration in Mind and Brain Science* (pp. 145–163). Oxford University Press.

Feyerabend, P. (1975). *Against method*. London: Verso.

Field, S. M., & Derksen, M. (2021). Experimenter as automaton; experimenter as human: exploring the position of the researcher in scientific research. *European Journal for Philosophy of Science*, *11*(1), 1–21.

Garey, M. R., & Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness* (1st Edition ed.). New York u.a: W. H. Freeman.

Gentner, D. (2019). Cognitive science is and should be pluralistic. *Topics in Cognitive Science*, *11*(4), 884-891.

Gierasimczuk, N. (2010). *Knowing one's limits: logical analysis of inductive inference* Unpublished doctoral dissertation. ILLC, Univ. of Amsterdam.

Gold, E. M. (1967). Language identification in the limit. *Information and control*, *10*(5), 447–474.

Gold, E. M. (1978). Complexity of automaton identification from given data. *Information and control*, *37*(3), 302–320.

Goldreich, O. (2010). *P, NP, and NP-completeness: The basics of complexity theory*. Cambridge Univ. Press.

Goodman, N. (1983). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*.

Hume, D. (1739). *A treatise of human nature*. Oxford: Oxford University Press.

Irvine, E. (2021). The role of replication studies in theory building. *Perspectives on Psychological Science*.

Jarecki, J. B., Tan, J. H., & Jenny, M. A. (2020). A framework for building cognitive process models. *Psychonomic Bulletin & Review*, *27*(6), 1218–1229.

Kelly, K. T. (1996). *The Logic of Reliable Inquiry*. New York: Oxford University Press.

Kelly, K. T., & Schulte, O. (1995). The computable testability of theories making uncomputable predictions. *Erkenntnis*, 29–66.

Kitcher, P. (1990). The division of cognitive labor. *The Journal of Philosophy*, *87*(1), 5–22.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kwisthout, J. (2011). Most probable explanations in bayesian networks: Complexity and tractability. *International Journal of Approximate Reasoning*, *52*(9), 1452–1469.

Lewis, H. R., & Papadimitriou, C. H. (1997). *Elements of the Theory of Computation*. Prentice Hall PTR.

Lipton, P. (2003). *Inference to the best explanation*. Routledge.

Love, B. C. (2020). Levels of biological plausibility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1815), 20190632.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., ... Vazire, S. (2019). Preregistration is hard, and worthwhile? *Trends in Cognitive Sciences*, *23*(10), 815–818.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).

O'Connor, C., & Bruner, J. (2019). Dynamics and diversity in epistemic communities. *Erkenntnis*, *84*(1), 101–119.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge university press.

Primiero, G. (2019). *On the Foundations of Computing*. Oxford University Press.

Quine, W. V. (1951). Main trends in recent philosophy: Two dogmas of empiricism. *The Philosophical Review*, 20–43.

Solomonoff, R. J. (1964). A formal theory of inductive inference. Parts I and II. *Information and Control*, *7*(1 and 2), 1–22 and 224–254.

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2020). Is preregistration worthwhile? *Trends in Cognitive Sciences*, *24*(2), 94–95.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–502.

Thagard, P. (2012). *The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change*. MIT Press.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, *22*(1), 1–24.

Thoma, J. (2015). The epistemic division of labor revisited. *Philosophy of Science*, *82*(3), 454–472.

van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories: A tutorial by dialogue. *Social Psychology*, *51*(5), 285–298.

van Rooij, I., Blokpoel, M., Kwisthout, J., & Wareham, T. (2019). *Cognition and Intractability: A Guide to Classical and Parameterized Complexity Analysis*. Cambridge Univ. Press.

van Rooij, I., Wright, C. D., & Wareham, T. (2012). Intractability and the use of heuristics in psychological explanations. *Synthese*, *187*, 471–487.

Varma, S. (2014). The subjective meaning of cognitive architecture: a Marrian analysis. *Frontiers in Psychology*, *5*, 440.

Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of science*, *76*(2), 225–252.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural computation*, *8*(7), 1341–1390.

Zollman, K. J. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, *72*(1), 17.