How Is Perception Tractable?

Abstract: Perception solves computationally demanding problems at lightning fast speed. It recovers sophisticated representations of the world from degraded inputs, often in a matter of milliseconds. Any theory of perception must be able to explain how this is possible; in other words, it must be able to explain perception's *computational tractability*. One of the few attempts to move toward such an explanation has been the information encapsulation hypothesis, which posits that perception can be fast because it keeps computational costs low by forgoing access to information stored in cognition. I argue that we have no compelling reason to believe that encapsulation explains (or even contributes to an explanation of) perceptual tractability, and much reason to doubt it. This is because there exist much deeper computational challenges for perception than information access, and these threaten to make the costs of access irrelevant. If this is right, it undermines a core computational motivation for encapsulation and sends us back to the drawing board for explanations of perceptual tractability.[1]

I. Introduction

Perception is hard. It is so hard that one of the main challenges of philosophy of cognitive science is to account for how perception, of the kind seen in people, is possible at all. But why is it so difficult? One reason is that perception solves problems with staggering computational requirements. It delivers reasonable solutions to these problems most of the time. And it does this all with very little of the resources central to computation: very little time, very little energy, very little data. No method in contemporary AI approaches these capabilities.[2] We can call this dramatic efficiency the *computational tractability* of human perception.

A theory of perception should explain how perception is computationally tractable. (Previous work in multiple traditions have defended tractability as a general constraint on mental processes and such arguments form the basis for research programs such as bounded rationality (Simon 1997), ecological rationality (Gigerenzer 2011), the tractable cognition thesis (Van Rooij 2008, Kwisthout 2011, 2018, Szymanik & Verbrugge 2018), and Bayesian resource rationality (Griffiths et al. 2015, Gershman et al. 2015, Icard 2018)). To date, however, few potential explanations have been offered. One notable exception is Information Encapsulation Hypothesis, which holds that perception is barred from accessing information stored in cognition. Proponents of the information encapsulation hypothesis often offer a computational motivation, suggesting that encapsulation helps account for the tractability of perception (Fodor 1983, Pylyshyn 1999,

[1] I am indebted to many people for help developing the ideas in this paper. For comments on (sometimes multiple) earlier drafts of this paper, I'd like to thank EJ Green, Jack Spencer, Alex Byrne, Laurie Paul, Ned Block, Agustín Rayo, Josh Tenenbaum, Bob Stalnaker, Kevin Dorst, and three anonymous referees for this journal. For help editing, I'd like to thank Madeline Medeiros Pereira. For discussions of these and related ideas, I'd like to thank Luke Hewitt, Jon Gauthier, Eric Mandelbaum, Johan Kwisthout, Scott Aaronson, and many others.

[2] For comparisons between human abilities and those of contemporary AI systems, see Lake et al. 2017, Kim, Ricci & Serre 2018, Marcus 2020, Firestone 2020, Jacob et al. 2021.

Mandelbaum 2017, Quilty-Dunn 2019).[3] Encapsulation, it is thought, explains (or partially explains) tractability by ensuring that perceptual processing does not incur the computational costs of search through large stores of information in cognition, as it would if perception were unencapsulated. Call this the Encapsulation Explanation of Tractability (EET).

In this paper, I argue that we have no positive reason to believe the EET, and many reasons to doubt it. Given what we know about the science of computational costs, information encapsulation seems to be the wrong kind of thing to explain the computational tractability of perception. In particular, encapsulation is ill-equipped to account for computational tractability because there exists a vastly larger problem for perceptual tractability than the cost of information access. I argue that, in light of the true landscape of computational costs inherent in perception, encapsulation can be neither necessary nor sufficient for tractability and is unlikely to even be a difference maker.[4]

If this is right, the implications are threefold. First, while it is still an open empirical question whether perception is informationally encapsulated from cognition, a core motivation for the thesis is cut adrift. This leaves the thesis more dependent on the weight of the psychophysical and neuroscientific evidence, bereft of a computational *raison d'être*. Since the empirical question is hotly debated (Macpherson 2011, Firestone & Scholl 2016, Lupyan 2017, Quilty-Dunn 2019, Green 2020), the loss of computational motivation matters a great deal to how we view the thesis. At stake in the encapsulation debate more broadly are issues of central importance to epistemology, such as whether perception can be treated as justificatory bedrock (Siegel 2012, 2017, Silins 2016, Jenkin 2020), and to philosophy more broadly, such as whether any distinction can be drawn between perception and cognition at all (Clark 2013) and how that distinction is to be spelled out if so (Phillips 2019, Green 2020).[5]

Second, our discussion places a strong constraint on future theories of perception. At the end of the day, we do not know how perception is computationally tractable, but a deeper understanding of the problem provides a better understanding of what a future solution must look like. I discuss constraints on a future theory of perceptual tractability in Section (VI).

---

[3] Tractability is not the only motivation for information encapsulation – encapsulation has also been offered as an explanation for striking psychophysical data, such as persistent illusions (see Muller-Lyer illusion).

[4] I'll consider that if encapsulation is any of these (necessary, sufficient, or a difference maker) then it is an explanation of perceptual tractability.

[5] Information encapsulation is one way in which perception might be modular, but there are others.

Finally, revisiting tractability arguments for information encapsulation has ramifications for theories of cognition more generally. If systems that are unencapsulated are *thereby* computationally intractable, then the traditional view of a unified mind post-perception is incompatible with the computational theory of mind; the view that mental processes are computational processes.[6] It would follow that either central cognition too must break down into parts that are encapsulated from one another (Tooby & Cosmides 1992, Pinker 1997, Carruthers 2004) or that the computational theory of mind must be abandoned (Fodor 2000). A re-evaluation of the connection between encapsulation and tractability will shed light on what is right, and what is wrong, with such arguments.

The paper proceeds as follows. Section (II) motivates the problem of computational tractability as it pertains to perception. Section (III) presents the solution offered by information encapsulation and the classical arguments for it. A formal definition of computational tractability as it is relevant to debates in the science of mind is developed in Section (IV). Sections (V) and (VI) argue that there exists a vastly larger problem of computational tractability than the one encapsulation was designed to solve, while Section (VII) argues that, in light of this, encapsulation can be neither necessary nor sufficient for tractability, and is unlikely to even be a difference maker. Some implications of this for the future of tractability arguments are presented in Section (VIII). Section (IX) concludes.

II. Why There is a Problem of Tractability

A venerable tradition in philosophy and psychology holds that perception is computationally tractable because it is informationally encapsulated from cognition (Fodor 1983, Tooby and Cosmides 1992, Pinker 1997, Pylyshyn 1999, Fodor 2000, Mandelbaum 2017, Quilty-Dunn 2019). In a moment we'll look at what information encapsulation is and how it is meant to address issues of tractability, but before evaluating potential answers, we should get clear on the question. Why should we think that perception has a computational tractability problem in the first place?

For the purposes of this paper, perception is the set of mental processes dedicated to gathering information by way of the sensory surfaces (such as the retina for vision or the cochlea for audition). This

[6] I.e. processes characterized by an abstract causal organization that mirrors the stages of a formal computational process (Chalmers 2011), in tandem with whatever relations to the environment are necessary to make some of those states representations (Fodor 1975).

includes the final stages of these processes, the perceptual outputs.[7] A good part of what perception does is solve *inverse inference problems*, in which latent causes are recovered or 'inferred' from their proximal effects. In the case of human perception, the latent causes are distal objects and their properties, and their proximal effects are their effects on the sensory surfaces, such as the retina, skin, or cochlea. In the particular case of vision, a set of properties including the shape, orientation, color, and distance of an object must be inferred from their joint effect – an image of colored light projected onto the retina. In nearly all real world cases of inverse inference, the proximal effects underdetermine the distal causes.

Inverse inference shows up everywhere in the mind, not just in vision. Audition performs inverse inference when it separates out particular voices or other auditory objects from an undifferentiated stream of vibrations, as when listening to someone talk in a crowded room. It is not limited to individual senses either. Perceptual inferences that recover the events associated with sounds take inputs from both audition and vision (a fact responsible for the ventriloquism effect, see Alais and Burr 2004), while the inferences that recover the shapes of objects take inputs from both vision and touch (Ernst and Banks 2002). Nor is it unique to perception. Cognition, by which I mean the set of mental processes of which reasoning and planning are paradigmatic examples, solves similar problems. When we infer that it rained from the fact that the ground is wet (when it could have been the sprinklers), that the neighbor is home from the fact that their car is outside (when they could have left on foot or bike), or the identity of a criminal from the evidence at a crime scene (which is consistent with any number of identities and scenarios), we are solving inverse inference problems. Other examples are less obviously causal, but are formally homologous, such as learning concepts from a finite set of examples, consistent with multiple hypotheses about their content (Feldman 2000, Xu and Tenenbaum 2007, Goodman 2008) or learning the theoretical relations that govern a novel domain (Gopnik et al. 2004, Tenenbaum et al. 2011, Ullman et al. 2012).

The fact that perception and cognition solve inverse inference problems is interesting because these problems are *hard*. They're hard enough that current methods for solving real-world inference problems either take a very, very long time to run (Sokal 1997, Park & Haran 2018) or huge amounts of time, energy, and data to

[7] For many authors, these outputs are synonymous with perceptual experience, see e.g. Firestone & Scholl 2016, p.1: 'There is a deep sense in which we all know what perception is because of our direct phenomenological acquaintance with percepts – the colors, shapes, and sizes (etc.) of the objects and surfaces that populate our visual experiences. Just imagine looking at an apple in a supermarket and appreciating its redness (as opposed, say, to its price). That is perception... Throughout this paper, we refer to visual processing simply as the mental activity that creates such sensations; we refer to percepts as the experiences themselves, and we use perception (and, less formally, seeing) to encompass both (typically unconscious) visual processing and the (conscious) percepts that result.'

train (Marcus 2020). The most recent work in AI illustrates these challenges. Training state-of-the-art language models (which infer likely completions from portions of sentences, for example, requires data sets on the orders of trillions of words (Brown et al. 2020) and days or weeks of computing time on hundreds or thousands of machines (Narayanan et al. 2021, Chowdhery et al. 2022). Contemporary vision models, which infer 3D-scene properties from images, are similarly compute intensive (e.g. Karpathy 2021).

In contrast, people solve inference problems quickly, cheaply, and with little training data (Lake et al. 2017, Marcus 2020). Why is the human mind so startlingly efficient? How is inference in the mind possible on the timescales that human-beings solve them? This is the first question of the tractability of the human mind. Call it the question of *absolute* tractability. Answering this question should tell us how computational systems could operate to solve inference problems in roughly the neighborhood of how long people take on those problems. Things that people solve in milliseconds should not take days of compute time. Things that take humans hours to learn should not take months. The question of absolute tractability applies equally well to both perception and cognition.

There is also a second question of tractability which is unique to perception. Even against the backdrop of the computational efficiency of cognition, perception stands out. While cognitive inference problems, such as concept learning, take many trials, encompassing seconds or minutes in the lab (Kemp et al. 2012) or hours or days in classroom and developmental settings (Carey 2009, Ullman 2012), perception solves its inference problems in record speed. For example, perceptual categorization of natural scenes on the basis of category (in this case, 'animal present' or 'animal absent') can be made within 150 milliseconds, as detected by EEG (a measure of the brain's electrical activity; Thorpe et al. 1996), while rapid eye movements or 'saccades,' which require motor planning as well as perceptual processing, can be made on the basis of similar categories in a few hundred milliseconds (Kirchener & Thorpe 2006). Changes in the neural decodability of stimulus information shows that by 350 milliseconds processing in visual areas has largely run its course, with perceptual outputs passed on to frontal, cognitive areas (Marti & Dehaene 2017).

Of course, speed alone is not impressive. It's easy to answer a problem quickly if one is willing to sacrifice performance. In the limit, problems can be answered randomly as quickly as one can roll an internal die. What is remarkable is perception's combination of speed *and performance*. Findings that support the optimality of perception (performance that reaches theoretical limits) are common in the field (e.g. Kording & Wolpert 2004, Ernst & Banks 2002, Weiss et al 2002, see Ma 2010 for a review). Other authors push back (see e.g. Rahnev and Denison 2016). Far less controversial is that perception's accomplishments are both impressive and unparalleled. It represents the world accurately enough that we get by in the myriad tasks we undertake and the diverse and open-ended environments in which we do them. Human beings rarely look at familiar objects and wonder what they are. We can pick out objects from a crowded visual field, recognize their distances and navigate to them, avoiding obstacles in the process. And we do this in all manner of circumstances: in various weather and lighting conditions, when viewed from different angles, and in novel surroundings. While contemporary machine learning systems can often beat human beings by a few percentage points in speeded classification tasks (Dodge and Karam 2017, Geirhos et al. 2018), human beings are unparalleled in their ability to recover 3D-scene geometry (Spelke and Kinzler 2007), object parthood (Green 2017), physical and relational properties (Wu et al. 2015, Hafri et al. 2013, Little & Firestone 2021), and the consequences of these for high-level features such as stability (Battaglia et al. 2013, Ullman et al. 2017, Hafri & Firestone 2021). Reproducing such accomplishments is the holy grail of computer vision.

There are then two distinct problems of the tractability of perception. The first, the problem of *absolute* tractability, is common to both perception and cognition. This is the problem of how either system is able to accomplish inverse inference on human-like timescales despite theoretical costs and engineered systems that suggest compute times well beyond this (much more on this to come). The other is the question of how perception manages to be so much more efficient (more tractable) than cognition, clocking in at speeds orders of magnitude faster than comparable processes in cognition. We can call this the *relative* tractability of perception (relative to cognition). Theories of perception must explain both the absolute and relative tractability of perception. This is a demanding requirement, but pursuing it vigorously is likely to be productive. Insofar as most theoretical frameworks for perception fail to account for tractability, insisting that a theory does so will help us cull the space of hypotheses as to how perception works.

III. The Encapsulation Explanation of Tractability

An architecture of perception must explain perception's impressive combination of speed and performance, both absolute and relative to cognition. Proponents of information encapsulation endorse the following explanation:

> **Encapsulation Explanation of Tractability (EET):** The computational tractability of perception is explained by the information encapsulation of perception from cognition.

The EET invokes the key concept of information encapsulation, but what exactly is this? Proponents of the thesis write:

> Looked at this way, the claim that input systems are informationally encapsulated is equivalent to the claim that the data that can bear on the confirmation of perceptual hypotheses includes, in the general case, considerably less than the organism may know. (Fodor 1983, p. 69)

> This target article … defends the position that an important part of visual perception … is prohibited from accessing relevant expectations, knowledge, and utilities in determining the function it computes – in other words, it is cognitively impenetrable. (Pylyshyn 1999, p.1)

> [This article focuses on]… traditional questions of whether visual perception is modular, encapsulated from the rest of cognition, and "cognitively (im)penetrable." At issue is the extent to which what and how we see is functionally independent from what and how we think, know, desire, act, and so forth. (Firestone & Scholl 2016, p. 2)

> With all this in the background, we can give a more precise characterization of encapsulation: System A is encapsulated from System B when A has a proprietary store of information that excludes information stored in B. (Quilty-Dunn 2019, p.3)

Each of these quotes seem to turn on some common idea, but there is also significant ambiguity. Fodor writes only that the information available to perception is 'considerably less' than is available to the entire organism. Pylyshyn prohibits access of the 'relevant' expectations, knowledge, and utilities, although it seems unlikely that he thought the irrelevant varieties of these could be accessed. Firestone and Scholl are interested in the 'extent' to which what and how we see is independent from what and how we think, know, and desire, while Quilty-Dunn endorses the generic, that 'system A's information store excludes information (Some of it? All of it?) stored in B.'[8]

While the specifics might be hazy, the gist is clear – the encapsulation of perception means that information in cognition is *verboten* for perception. Fodor thinks that the information available to perception is

[8] In context, it's clear that the Quilty-Dunn quote should receive the universal reading.

considerably less than the organism may know because none of the information in cognition is available to it. Pylyshyn points out that the relevant expectations, knowledge, and utilities are prohibited because all of the expectations, knowledge, and utilities are prohibited.[9] Firestone and Scholl are interested in the extent to which what and how we see is independent from what and how we think, know, and desire because they want to defend the view that perceptual processing is not influenced by any of these. Information encapsulation then is a relational property. One system is informationally encapsulated from another when the first is barred from accessing the information in the second. In this case, the relevant kind of information encapsulation is the encapsulation of perception relative to cognition.

The strongest version of the thesis is that none of the information in cognition is accessible to perception. This universal reading is reasonably natural and satisfies the conditions given by each of the quotes. It is further motivated by the content of the papers and chapters from which these quotes are drawn.[10] It is also possible that the universal reading is too strong. Perhaps it is enough if *most* of the information in cognition is barred from access by perception. If this were an empirical paper, with the aim of providing a counterexample to encapsulation, a lot would turn on whether encapsulation theorists are committed to the universal thesis. As it stands, the project of this paper is to show that information encapsulation makes at best a negligible contribution to an explanation of the computational tractability of perception. For this, the strongest version of the thesis will do just fine. If a prohibition on all of the information in cognition is not enough to meaningfully impact tractability, a weaker version of the constraint is unlikely to fare better.

The next question is, *how* is information encapsulation meant to explain tractability? Proponents of the EET write:

> … speed is purchased for input systems by permitting them to ignore lots of the facts. (Fodor 1983, p. 70)

[9] At least for early vision.

[10] In each case, what follows are arguments challenging a swath of psychological results that have been taken to evince the the accessibility of information in cognition on perceptual processing, either because such effects are consistent with an explanation citing only information available in perception (Fodor 1983, Pylyshyn 1999), because the effects can be reproduced where the theory of cognitive penetration predicts they should not be (Firestone and Scholl 2016), or because effects that might look like cognitive penetration are in fact mediated by attention, rather than access (Quilty-Dunn 2019).

> One of the reasons theorists have been drawn to modularity theory is its evolutionary rationale.... Roughly, the intuition is that during panther identification what really matters is accomplishing such identification quickly... Searching through everything we know about panthers in order to make an identification would be extremely time consuming. (Mandelbaum 2017, p. 10)

> How are perceptual processes computationally tractable? ... If the processes that solved these problems had to sift through all information stored in central cognition, they would face an unwieldy computational burden... If instead perceptual processes are encapsulated, then they need only check input against their proprietary stores of information ... Encapsulation can therefore provide a unified account of perceptual processes as computationally tractable operations that occur outside of central cognition. (Quilty-Dunn 2019, p. 5)

The thought seems to be that retrieving information takes time, retrieving information from larger stores takes longer, and retrieving the relevant information from a store of information as large as cognition would take *too* long. By foregoing this expense, however, perception can be accomplished tractably.

We can call this basic idea the *Haystack Idea*. Finding a needle in a haystack is a hard problem (hard enough that it has become an idiom for difficulty) and finding relevant information in a huge store of information poses a similar problem. Moreover, needle-in-a-haystack problems get harder as the haystack gets larger.

If we run with this idea for just a moment, we can also get a sense for how the problem 'scales,' or gets harder, as the number of inputs changes. Intuitively, every additional entry makes the problem a little bit harder in expectation. For concreteness, we can think of the set of possibly relevant entries as a list. Entries on this list are information in whatever format one thinks that the mind represents it – this could be a list of beliefs written in the language of thought (Fodor 1975, Goodman et al. 2015), natural constraints (Marr 1982), or parameter values of graphical models (Danks 2014), to name just a few possibilities. Under the pessimistic assumption that this list is unordered – that is, that we do not know in advance where relevant information is to be found (Fodor 1983, 2000) – the expected number of steps needed to find relevant information grows linearly with the number of entries.[11] If the haystack is large enough, search will take too long. Under these conditions, an artificial limit in the size of the haystack, the encapsulation of perception relative to cognition, could explain the tractability of perception. This then is the key idea motivating the EET. The EET holds that the tractability of perception can be explained by avoiding the linear costs of search through the information stored in cognition.[12]

A few clarifications about the EET are in order before we continue. These concern the kind of tractability (relative or absolute) that the EET is meant to explain, the kind of explanation the EET is meant to offer (whether a sufficient condition for tractability, a necessary condition, or a difference maker), and the empirical assumptions the EET requires to get off the ground. I'll look at each of these in turn.

First there is the question of the target of explanation. We noted above that there are two questions of tractability. One asks how perception could be tractable relative to cognition – that is, why perception is orders of magnitude faster than cognition, despite solving mathematically similar problems. The other wonders how perception could be tractable in absolute terms, which is to say, how perceptual processing can happen on roughly human timescales. From what we've seen so far, it might seem that the EET is best suited as an explanation of the *relative* tractability of perception. This is a modest version of the thesis. On this interpretation, the EET is silent on the question of how cognition and perception are accomplished with merely human levels of compute.[13] Instead it merely aims to explain the speed of perception relative to cognition (the difference between, say, minutes for thought and milliseconds for seeing).

Some proponents of EET likely understand the thesis in its modest version. This is just as well, as the immodest version of the thesis leads to some strange consequences. For example, if the price of *absolute* tractability is forgoing access to information stores on the scale of those that exist in cognition, then it follows that cognition itself must be divided into parts, none of which exceeds that critical threshold, or that cognition must be computationally intractable. Interestingly, *both* of these consequences have been endorsed by theorists working in this tradition. Massive modularists, such as Tooby and Cosmides (1992), Pinker (1997), and

---

[11] If sampling randomly over the unordered list, the growth in expectation of a geometric distribution with probability of success i/n, where n is the number of entries where the information can be found and i is a constant. Deterministic search over an unordered list (i.e. a list where order is independent of relevance) is equivalent to sampling randomly without replacement, an unnamed distribution which also scales linearly in expectation as n grows. The costs are also linear in the worst case, when all the information must be accessed, in which case the costs increase at a rate of exactly N steps per entry, where N is the number of steps required for access.

[12] Other operations, other than search, scale non-linearly, either in the number of entries or in other parameters. We'll do a deep dive into such costs in the sections to come. For our purposes now, the essential takeaway is that the costs of search scale at worst linearly, even under strong pessimistic assumptions about the efficiency of that search.

[13] Rather than the industrial levels of compute required by today's AI (see Section II) or the astronomical levels of compute suggested by theoretical analyses (more on this in a moment).

Carruthers (2007), give up on the idea of a unified central cognition, citing tractability arguments, among others.[14] These theorists prefer a view of cognition on which the mind is really a bundle of independent cognitive entities, each working on certain ecologically salient problems. Opting for the other horn, Fodor himself held that the integrated nature of central cognitive processing is undeniable and argued on these grounds that a computational theory of mind could never include central cognition![15][16]

How should we understand the EET then? As an explanation of relative or of absolute tractability? For the purposes of this paper, we won't ask the proponent of the EET to commit one way or the other. The argument developed below will show that information encapsulation is not the right place to look for an explanation of either kind of tractability.

Next, there is the question of what kind of explanation the EET is meant to offer. There are a few options here. One could hold that encapsulation is sufficient for tractability: that is, that perception must be tractable if it is encapsulated. Or that encapsulation is necessary for tractability: that perception could not be tractable unless encapsulated. Finally, a weaker version of the EET might grant that encapsulation is neither sufficient nor necessary for tractability, but maintain that encapsulation is nevertheless a difference-maker: that is, that perception would not be tractable were it not encapsulated. (This is different from either necessity or sufficiency. For example, striking a match is not sufficient for lighting a match, since there must be oxygen and the room. Nor is it necessary, since a match can be lit by other means. It is, nevertheless, a difference maker – holding fixed all else about the system, the match would not have been lit but for the striking.) If encapsulation explains tractability, then absent systematic overdetermination, it must at least be a difference maker. Here again, I won't try to pin down exactly which of these versions of the EET proponents have in mind. Instead, I'll argue against all three versions of the thesis. That is, I will argue that encapsulation is neither sufficient nor necessary for tractability, and that there is no positive reason to believe it is even a difference-maker.

[14] Carruthers (2007, p. 44-52) offers the clearest defense of massive modularity on tractability grounds. See also Tooby & Cosmides (1992, p. 106).

[15] Fodor writes, "Indeed, I am inclined to think that, sooner or later, we will *all* have to give up on the Turing story [of computation] as a general account of how the mind works…" (p. 47). Why? Because "…the computational theory of mental processes doesn't work for abductive inferences" (p. 41). This means that "… a cognitive science that provides some insight into the part of the mind that isn't modular may well have to be different, root and branch, from the kind of syntactical account that Turing's insights inspired." (2000, p. 99)

[16] Examples like these illustrate that some theorists clearly have an immodest version of the EET in mind, but not all versions of the EET lead to this dilemma. If the EET is meant to explain only the relative tractability of perception, then no such conclusions about cognition follow.

Finally, a few words about the empirical assumptions that the EET requires in order to get off the ground, which I'll be granting for the sake of argument. These assumptions fall into two categories. One set of empirical assumptions has to do with the distribution of information between perception and cognition. If the encapsulation of perception from cognition is meant to explain the tractability of perception in any of the senses discussed above, then such an explanation turns on contingent facts about just how much information is stored in each. If there is too much information stored in perception, for example, then perception will be intractable regardless of whether it is encapsulated (and so encapsulation cannot be sufficient for tractability). Conversely, if there is too little information in cognition, then eschewing access to such information will be neither necessary for tractability, nor a difference maker. The interest of the thesis therefore depends on facts about the relative amount of information in perception and cognition.

In what follows, I'll grant the encapsulation theorist the empirical facts that they seem to believe: that perception's proprietary store of information is small enough to be tractably searched in the time it takes for perceptual processing to unfold, and that cognition's store is meaningfully larger, such that searching cognition would represent a significant multiplier on the work involved in searching perception alone. (My main interest in this paper will not be in challenging any of these facts, but rather in taking issue with the underlying view of tractability that makes such facts relevant.)

The other set of empirical assumptions required to warrant interest in the EET has to do with the connection between information access and information search. Encapsulation bars information access, and the EET holds that foregoing such access explains tractability by keeping computational costs low. Strictly speaking however, information access costs hardly anything at all – all the relevant costs are the costs of search. To put the idea bluntly: finding a needle in a haystack can be challenging, but if someone gives you a needle from a haystack, receiving it is not difficult. Why couldn't cognition simply send the relevant information for perceptual processing to prime perception in the next moment, obviating the need for an expensive search on perception's part? Cognition could, for example, send a relevant color memory (Hansen et al. 2006, Machpherson 2012) or expectation about some other low-level feature (Kok et al. 2012). While this would be a violation of encapsulation, it wouldn't require anything like a perception-initiated, real-time, or full-scale search through cognition, and wouldn't require anything that any party to the debate currently believes to be intractable (no one denies that people can recall the approximate colors of objects from long-term memory or notice a pattern in the features of serially presented stimuli!). In such a case, the computational costs of a violation of encapsulation would be near-zero.

To connect the negligible costs of access to the more considerable costs of search requires some argument. Maybe evolution opted to prevent all cognitive influences, including ones that are obviously cheap, in order to avoid the costly ones? Such a scenario would be plausible if we assumed that evolution faced the choice between either barring all cognitive influences or barring none, but this idea rests on a strangely dichotomous view of the computational options available. After all, there are many ways in which access could be consistent with non-exhaustive search (some of which are discussed in Section VII) and no a priori reason to think such intermediate solutions are inaccessible to evolution. Be that as it may, we will assume that there is some argument of this type available to the proponent of the EET, as we can make sense of the view only if information access can be wedded to the computational costs of search.

We now have a sense of the breadth of versions of the EET and the empirical assumptions on which the plausibility of the EET depends. In the next section, we'll analyze the concept of computational tractability at work in the EET.

IV. Tractability as an Empirical Bound

We can begin with some points of agreement between all parties. If the mind is computational, then it has some basic operations for manipulating information.[17] These operations could be manipulations of explicit symbols according to rules, as in traditional computers, or the transformation of large vectors of inputs by matrix multiplication, as in contemporary neural networks, or something else besides. Because these operations are implemented in a physical substrate, each instance of an operation takes up some fixed, finite amount of time. It follows that doing too many such operations will take too long; i.e. will render a computation *intractable*. This line of reasoning gives us a simple account of computational tractability as it applies to the mind, which is common to proponents of EET and their detractors.

Tractability: A computational procedure is tractable when it can be completed in fewer than K steps.

A few clarifications are in order. A computational procedure is a finite set of instructions and basic computational operations which define a series of applications of those operations for each instance of a given class of inputs, delivering an output. Crucially, computational procedures must function without recourse to anything but their inputs, instructions, and basic operations (see Turing 1936). A computational procedure

[17] See Section I, p. 5 and footnote 6, for discussion of the computational theory of mind.

may be *branching* in the sense that it doesn't have to execute the same series of operations every time. It can treat different inputs differently (say, running a distinct series of operations for odd numbered inputs as for even), and could even be stochastic, making random choices at predefined points in its execution.

Computational procedures are the only way we know of to solve computational problems. A computational *problem* is a set of inputs, a set of outputs, a set of ordinal or metric structures over those outputs, and a mapping from inputs to a given structure over outputs. The metric or ordinal structure over the set of outputs reflects the fact that answers to computational problems are not always right or wrong, but are often better or worse than one another. Better procedures are those that deliver better performance on a problem.

A computational problem can also be tractable or intractable. A computational problem is intractable, relative to a performance criterion, when no procedure can tractably solve that problem to that performance specification. This bit of relativism is necessary for a meaningful notion of tractability, as there are few limits on how quickly an answer can be computed in the absence of non-trivial criteria for how good an answer it has to be (see Section II). Criteria may include how close one is to the right answer, how often one gives the right answer, the class or proportion of problem instances for which one gives the right answer, or any combination of these.[18] The performance criterion relevant for our purposes is that of *human-level* performance. To explain how, say, visual inference is tractable, is to explain how it could be computed to human-level performance in fewer than K steps.

Finally, we might wonder, what is K? For our purposes we can imagine that, for a particular problem and performance criterion, K is some fixed number, determined by how long it takes a human being to solve that problem to that performance criterion. For a given problem and performance criterion, there are many things that might affect how big K is. One is the speed of the relevant basic operations in the human brain. Faster operations permit a higher K. Another is parallelization. Some problems have parts that can be solved by parallel trains of operations, increasing the number of operations that can be packed into a unit time. The speed of basic operations in the brain and the extent to which parallelization is employed are both questions outside the scope of this paper. In light of substantial uncertainty about these values we should err on the side of liberality when setting K, so as not to prematurely eliminate hypotheses about the mind that we can't be sure are intractable. In

[18] An acceptable criterion for performance for visual estimation of distance for example, could be that the visual system delivers an answer within 20% of the true value, 90% of the time, when presented with an object in good light.

other words, we should allow that K for many human perceptual processes may be quite large. It will not, however, be *astronomically* large.

What counts as astronomically large? We can gain something of a foothold on this concept by starting with the capacities of today's supercomputers. Today's fastest supercomputers perform on the order of 10^{17} operations per second. To do this, they run thousands of processors, occupy whole complexes, and consume vast amounts of power. For the purposes of this paper, we'll say that anything that would take one billion supercomputers one billion seconds (~115 days) to compute (that is, $>10^{30}$ operations) is 'astronomical.' Trivial as it sounds, we will see that the requirement that perceptual processing not require astronomically many steps will turn out to be a constraint with some teeth.

Taken together, an explanation of the tractability of perception is an explanation of how perception is accomplished without astronomical costs. Over the next two sections, however, we'll see that plausible assumptions about the costs of perceptual inference actually do entail astronomical costs. To explain the tractability of perception then, a theory must explain how these assumptions can be denied and these costs avoided. This will give us a positive framework for thinking about tractability. Finally, we'll see in Section (VII) that information encapsulation is ill-equipped to contribute meaningfully to an explanation of tractability in light of all this.

V. Inference & Scaling Behavior

A theory of perceptual tractability must explain how perception is accomplished, relative to human-level performance, by some computational procedure that takes less than astronomically many steps. If we call the number of steps needed to solve a problem to the relevant performance criterion M, then a perceptual inference is tractable when $M < K$. But what properties of a problem contribute the most to M for the computational procedures that solve it? And, in particular, what properties put us at risk of astronomically large M?

Theoretical computer science can be a source of insight here. One branch of theoretical computer science, Computational Complexity Theory (CCT), reasons about computational costs through the lens of how M grows, or 'scales,' with different properties of a computational problem, including most famously the number of inputs to a problem, but with theoretical extensions to include considerations of performance criteria and distributions over inputs. Co-opting some of the core concepts of this field will help us better understand our own notion of tractability. (For more detailed introductions to CCT then I can provide here, see Sipser 2012, Arora and Barak 2007, or Goldreich 2008.)

CCT and Scaling Behavior

CCT taxonomizes computational problems according to the functional form of the 'scaling behavior' of a problem on the number of inputs. The scaling behavior of a computational *procedure* is the relationship between the number of inputs and the number of steps the procedure goes through for those inputs, while the scaling behavior of a *problem* is the behavior of the most efficient procedure for solving it. To get a feel for how scaling behaviors differ between problems, imagine attempting to plan a wedding given a guest list. If you want to know whether you'll need chairs or not, all you have to do is check whether the list is non-empty. This is a constant time operation; and this takes equally many steps no matter how long the list is. If you want to know how many chairs you'll need, by contrast, then you need to count the number of names on the list. This is a linear time operation; it requires a number of steps that is a linear function of the length of the list. Finally, if you want to know what seating arrangement will maximize the well-being of your guests, allowing old friends to catch up, kindling new romances, and avoiding explosive tiffs, you'll need to consider every way your seating chart could be arranged. This is an exponential time operation. Such operations tend to be sticking points in our lives, as in the lives of computers. No one complains about having to count the number of guests on a list, but planning the seating can be a nightmare.

CCT treats this difference between exponential and sub-exponential scaling as a difference in kind rather than degree. It does this because, for moderately sized inputs and beyond, the contrast between exponential and sub-exponential scaling often separates operations that can be feasibly computed, even at significant cost, from those that cannot. For example, in the wedding case above, if your guest list contains 90 people, then checking whether you'll need seating takes 1 step, while counting how many chairs takes 90. If you can set up at most 10 tables of variable size for your guests, then finding the optimal seating arrangement requires on the order of 10^90 steps, at least one for each of the unique possibilities that must be considered. 10^90 is a big number; it's more than the number of atoms in the known universe. For the purposes of practical computation, it might as well be infinite. That's why CCT treats this difference in degree as an effective difference in kind.

A few clarifications. First off, not all inputs to a computational problem contribute equally to the cost of solving it.[19] We'll discuss this at length in the case of perceptual inference in the next section. Second, CCT makes strong assumptions about the performance criteria relative to which costs are assessed – most famously requiring guaranteed performance on all (and therefore the most difficult) problem instances — and this limits its relevance to our project.[20] Relatedly, CCT doesn't model aspects of problem structure that might make certain problem instances easier or harder. Where particular classes of problem instances have additional structure, that structure can sometimes be exploited to make a problem in that class easier than the complexity of its super-class would suggest.[21] When it comes to the project of understanding computational complexity as it applies to theories of the mind, we'll take what we can use from CCT and leave what we can't.

The key thing we will keep is CCT's focus on scaling behavior. This simple idea is both deep in what it reveals about the nature of computational problems and crucial for the task at hand. To be fit for our purposes, however, the concept of scaling behavior drawn from CCT will have to be both broadened (to include more diverse performance criteria) and refined (so as to be applied to classes of instances that have exploitable structure). Our CCT-inspired examination of scaling behavior will give us a place to start in examining what properties of perceptual inference problems entail which computational costs.

The argument will proceed as follows. We'll begin by establishing some general facts about the way that the *hypothesis space* of an inference problem grows as a function of the *dimensionality* of that problem. What we mean by these words will be made clear in due course. We'll find that the hypothesis space grows exponentially as a function of dimensionality. Under some simple starting assumptions, this exponential growth translates into exponential growth in the costs of computing inference. We'll then see that these assumptions can be substantially weakened, leaving the main result intact. In the following section, I'll argue that this exponential growth in the costs of perceptual inference, combined with the actual dimensionality of real life perceptual inference problems, suggest astronomical costs for perceptual inference. These costs dominate anything else in

---

[19] For example, the computational costs of determining whether a formula in propositional logic is satisfiable is exponential in the number of literals that appear in the formula, but not exponential in the length of the formula. More detail on this area of research, known as Parameterized Computational Complexity Theory, can be found in Downey & Fellows (2013) and Flum & Grohe (2006). See Kwisthout (2011; 2018) for an overview of key results related to cognitive science.

[20] Other branches of CCT look at average performance assuming simple distributions over inputs. This too is unlikely to be the kind of performance criterion most relevant to a computational theory of the mind, since ecologically realistic distributions over inputs are often not simple and generally lack closed form expression.

[21] Backtracking Satisfiability (or 'SAT') solvers (e.g. Davis & Putnam 1960, Davis et al 1962) are a classic example of a strategy that exploits local problem structure to find a solution more quickly. Since a minority of SAT cases exhibit global structure that frustrates such strategies, SAT exhibits a disconnect between theoretical intractability and computational procedures that are tractable for most purposes.

perceptual processing. One consequence of this is that explaining the tractability of perception requires explaining how some of these assumptions can be credibly rejected so as to avoid astronomical costs. Another is that the massive computational costs intrinsic to perceptual processing undermine any intimate explanatory connection between information encapsulation and perceptual tractability.

Scaling Behavior of Inference

The first concepts we'll need are those of a *hypothesis space* and the *dimensions* that define it. Solving an inference problem requires finding one or more good hypotheses about how the world might be from the set of all the ways the world could be, at least by the lights of that inference problem. In the wedding planning example, 'all the ways the world could be' includes all the ways that people at your wedding could be seated. (The problem is 'blind' to many other ways the world could be, such as how the astronauts in the International Space Station might be seated.) Hypotheses, or candidate solutions to the inference problem, differ from one another in their assignments of values to variables, such as people to tables. These variables can be thought of as the 'dimensions' of a space (the 'hypothesis space'), the values as coordinates along those dimensions, and hypotheses as unique points in the space.[22] In the wedding planning example, each attendee is a 'dimension' which must be assigned a value, in other words, a table. In the case of visual, perceptual inference, which we'll get to shortly, the hypothesis space is given by all the ways the objects in a scene could be – their colors, shapes, locations, etc.

One thing to notice is that the dimensionality of an inference problem (which dictates what hypotheses can be represented) comes apart from the information we bring to bear in solving that problem (what is known about those hypotheses). Adding a new dimension – say, kind of chair – allows the system to formulate new hypotheses (should I seat Veronica and Ezra in bean bags?), while information changes the assessment of quality of various hypotheses (I might know that Matthias would not like being sat at the kids' table). The distinction here is not idle. While being able to represent a dimension offers a natural way to represent information about

[22] Here I am eliding the question of whether to think of the hypothesis space as a semantic feature of the problem (ways the world could be) or as a syntactic feature of the representation of the problem (ways the world could be from the perspective of a procedure). In the absence of meaningless or synonymous expressions and assuming that all relevant hypotheses are expressible, there will be a 1-to-1 correspondence between hypotheses in the syntactic sense and in the semantic sense. Assuming that a problem can be fully represented then, deviation from this 1-to-1 correspondence comes when there are more syntactic hypotheses than there are genuine possibilities. Since a computational procedure can only operate over syntactic hypotheses, such deviations create additional costs. On the assumption that all semantic differences can be represented then, we can treat the semantic dimensionality of a problem as a lower bound on the cost-driving, syntactic dimensionality of a computational procedure for solving it.

that dimension, a system can also represent a dimension without having any information about it.[23] Similarly, a system can make use of information that is encoded in dimensions it does not represent. An example of the first would be if the visual system could represent colors and object categories, but was encapsulated from relevant information in cognition about the colors of known objects. In this case, vision possesses the dimension of color, but lacks information about it. An example of the second would be if the visual system could represent color and face identities, but not party affiliation. Cognition, for its part, might know that a particular person is a republican and that republicans are likely to wear red ties. We can imagine that cognition sends a visual expectation down to vision about the color of a tie in response to a perceptual output recognizing the face. In this case, vision would possess information about the dependency between identity and tie color, while lacking the dimension of party affiliation that introduces the connection. This distinction, between information and dimensions, will be critical in what's to come since the costs of inference are sensitive to the two in very different ways.

How exactly are the costs of inference related to dimensionality? To ease into this, think first about how the set of hypotheses grows as new dimensions are added to the space. When we add a new dimension, each possible value of the new dimension combines with every previously complete hypothesis to deliver a new set of unique hypotheses. So, for example, if we add 'kind of chair' to our wedding planning problem, then where we previously had a single complete hypothesis (a total assignment of people to tables), we now have a set of hypotheses; every possible combination of assignments of people to types of chairs, consistent with a table assignment. Just like adding a new dimension to a real coordinate space, this produces exponential growth in the set of hypotheses.

So our set of hypotheses grows exponentially as the dimensions of the problem are increased. But how is this tied to the actual costs of performing inference? Some simple assumptions will deliver the result that exponential growth in the hypothesis space produces exponential growth in the costs of inference.[24] I'll first present these assumptions and then look at ways they might be weakened.

[23] If some information is necessary for concept possession, then read this as 'without any further information than is necessary for representing the dimension'.

[24] The thought here is that the costs of inference are an exponential function of the dimensionality of the problem (I'll show later that this, combined with the (possibly fixed) dimensionality of perceptual inference problems is sufficient to put the costs of inference in astronomical territory). The talk of exponential 'growth' in costs is merely meant to bring out this functional relationship. It will not be important to my argument whether the dimensions of perceptual inference can continue to be increased (in fact, they may be architecturally barred from doing so for just this reason; See Section VIII).

First, consider performance criteria. We saw that computing inference requires finding 'good' hypotheses from within an exponentially growing hypothesis space, where the goodness of a hypothesis consists in its probability, plausibility, or explanatory import. For the moment, take finding a 'good' hypothesis to mean finding the 'best' hypothesis. Next, assume that it costs at least one computational step to evaluate a hypothesis and only one hypothesis can be evaluated at a time.[25] Finally, assume that nothing is known beforehand about the relative or absolute distribution of good hypotheses throughout the space. That is, the only way to find out whether a hypothesis is any good is to evaluate its plausibility relative to a prior and the data.

When these assumptions are met, the computational costs of doing inference grow linearly with the number of hypotheses and therefore exponentially with the number of dimensions defining the hypothesis space. This is because hypotheses must be evaluated in order to determine their performance, and must be evaluated in some order that is independent of the performance of the hypotheses (since having access to an order that privileges better hypotheses would violate the assumption that nothing about hypothesis performance is known beforehand). This entails that the number of hypotheses that must be evaluated grows linearly in expectation with the number of hypotheses in the space, and therefore exponentially in the dimensionality. Note that this holds whether we are sampling randomly (with or without replacement[26]) or evaluating hypotheses in a predetermined order (which, since it cannot be relied on to privilege the best hypotheses in the general case, might as well be a random order). Finally, since the number of hypotheses that must be evaluated in expectation grows exponentially as a function of the dimensionality, and since the costs of evaluating a hypothesis are constant, the costs of evaluation grow exponentially as well.

This line of reasoning establishes exponential growth under these assumptions, but some may find the assumptions troubling. The performance criterion is a particular sticking point. While many have argued that human perception is optimal, in the sense of finding the best solutions to its inference problems (Ernst & Banks 2002, Weiss et al. 2002, see Ma et al. 2010 for a review), others have argued against this perspective (e.g. Rahnev

[25] We can actually get by with a much weaker assumption, i.e. the assumption that there are no exponential speed-ups in the number of hypotheses that can be evaluated at a time (either as a function of the amount of time spent reasoning or the number of dimensions in the hypothesis space). That is, we can get by with the assumption that there can be no finite equivalent of super-tasking: evaluating one hypothesis in c steps, the next in ½c steps, the next in ¼c steps, and so on. I use the stronger assumption that evaluating a hypothesis costs one computational step because it will considerably simplify the presentation of the argument in the following section.

[26] Assuming, for simplicity, a uniform distribution over hypotheses, the expected number of hypotheses sampled before finding the best corresponds to the Geometric distribution with exponentially decreasing probability of success as dimensions are added. Sampling without replacement yields another distribution that also grows linearly in expectation as the number of non-best hypotheses grows, and therefore exponentially in dimensionality.

& Dennison 2018). We can, however, weaken the performance criterion in reasonable ways while maintaining the main result. Imagine, for example, that instead of finding the best hypothesis for a given problem, human perception finds hypotheses that are merely 'good enough', in the sense that they are close enough in value to the best hypothesis along each of the dimensions of the problem. In this case, we might count as a satisfactory answer any hypothesis within 3% of the value of the best hypothesis along each of the relevant dimensions. This has the effect of turning a solution set from a point to a contiguous region in the hypothesis space. Such relaxations would certainly make these problems easier to solve, but they do not resolve the more fundamental issue of exponential scaling. To see this, imagine solving an inference problem to this 'good enough' standard of performance. Even in this case, the proportion of hypotheses meeting this criterion shrinks exponentially as the dimensionality of the space increases. For one dimension, 3% of samples will meet this criterion. But for 3 dimensions, that proportion is 0.0027%, for 6 it's 7.3 x 10^-8 %, and so on.[27] Here, as above, the proportion of good hypotheses becomes vanishingly small, and reasonable assumptions about the costs of evaluation will entail astronomical costs for finding those hypotheses. (This example also illustrates how exponential scaling generalizes to continuous hypothesis spaces, where in the continuous case, as in the discrete case, the proportion of the measures of the solution set and the problem set shrinks exponentially.[28])[29]

There are, of course, many ways to weaken the performance criteria, and we've only looked at one. It may be that some of these ways avoid exponential growth in the costs of inference while still delivering human-level performance. This is, however, not where I'd put my money. Human performance on perceptual inference tasks is excellent (see Section II). It seems for this reason that weakening the performance criteria to such an extent that hypotheses that meet those criteria will be easy to come by, even in astronomically large hypothesis spaces, is a non-starter. Instead, we'll have to ask which of our other assumptions can be given up, in particular the assumption that nothing is known in advance about the distribution of promising hypotheses. This will be a focus of a later section (Section VII). For the time being, we need to show that this theoretical

---

[27] That proportion is given by the equation (3/100)^n.

[28] The challenges of exponential scaling are also robust to reparameterization. While in either discrete or continuous cases one could always map the hypotheses from N dimensions onto a single dimension, such a trick would change neither the relative cardinalities of the solution set to the problem set in the discrete case nor the proportion of the measures in the continuous case. (In general, a syntactic representation of the problem that differs from the problem's intrinsic dimensionality can make the solving the problem more difficult, but it cannot reliably make it easier. See footnote 22.)

[29] Note that the astronomical costs of inference hold even if, as is believed, the brain is massively parallel (see Section III, p. 30). Since parallelization can deliver at best a factor of N speed-up, where N is the number of parallel processes, the exponential increase of costs is maintained regardless. Unless the number of parallel processes is itself astronomical (much greater than the estimated 10^7 neurons in the brain), parallelization won't make a difference in the analyzes to come.

result of exponential scaling actually is sufficient to push the costs of perceptual inference into astronomical territory when certain empirical facts about the dimensionality of perceptual inference are considered.

VI. Dimensionality of Perceptual Inference

I've argued that, under reasonable assumptions, the costs of inference scale exponentially in the dimensionality of the problem. But what does all this mean for the tractability of real world perceptual inference? To know what conclusions we should draw requires developing a rough idea of the dimensionality of perceptual inference and the proportion of perceptual hypotheses that satisfy human-like performance. I'll argue that conservative assumptions about dimensionality, and liberal assumptions about the proportion of hypotheses that satisfy human-like performance, combined with our assumptions so far, entail astronomical costs for perceptual inference. I'll make this case by presenting a toy visual inference problem, involving just a few of the many dimensions that vision represents.

A Toy Inference Problem

Consider a scenario in which I open my eyes to see a simple scene of static objects. Each object has a color, a lighting condition, a location in three dimensions, and a shape. We can set some numbers to these possibilities. Perhaps there are a million (10^6) colors we can see,[30] another million (10^6) ways the lighting could be (Tokunaga and Logvinenko 2010), perhaps a billion possible locations (10^9), and another billion (10^9) possible shapes.

These are conservative figures. Stipulating a billion possible locations amounts to assuming that there are a mere 1000 just noticeable differences in location across each of 3 dimensions – a modest estimate of human spatial acuity.[31] In the case of shapes, a mere billion discriminable possibilities across all the configurations of shapes and sizes perceptible to human beings is a gross underestimate. Even with such conservative numbers, however, the different combinations deliver 10^30 ways an object could be. If there are 3 objects in a scene, the

[30] Estimates range from 1-10 million.

[31] Just Noticeable Differences (JNDs) are the smallest differences that provoke above chance discrimination. Here I'm assuming for the sake of simplicity that there are an equal number of JNDs across each dimension. Distortions in visual space may mean that this is not quite right (Green & Rabin 2019). Note that the connection between perceptual hypotheses (distinct internal representations) and discrimination is not direct – distinct representational states are the competence to discrimination's performance. Discrimination is accomplished by mapping equivalence classes of stimuli to distinct representational states. Distinct representational states can, however, exist without showing up in discrimination, say if insufficient light, damage to the retina, or other peripheral constraints impair performance. Discrimination then places a lower bound on the number of distinct representational states; the value relevant for inference.

number of possible scenes is 10^90. Here again, this number is greater than the number of atoms in the known universe. For practical purposes, it may as well be infinite.

We saw earlier that loosening up the performance criteria within reason does not change the exponential decrease in the proportion of viable solutions. But where do these considerations get us in the case of perceptual inference? We already assumed that the divisions were coarse-grained (with just 1,000 JNDs in location along each spatial dimension). But we can go further. Let's say that any hypothesis is acceptable so long as it falls within a range of 3% of the best hypothesis, along each dimension in the hypothesis space. Assuming that color, lighting color, location, and shape each involve three dimensions, the proportion of hypotheses satisfying this condition for a scene with three objects would be on the order of 1 in 10^54 – still well within astronomical territory.[32]

In setting this up I have said nothing of numerous other dimensions represented in vision, including low-level dimensions such as edges, as well as many high-level contents, such as motion (Weiss et al. 2002),[33] object identity (Quilty-Dunn 2019), causality and animacy (Scholl & Tremoulet 2000), or hierarchical part structure (Green 2017). I have also neglected dimensions from other modalities which participate in inference in cross-modal perception (Green 2021) and cue integration (e.g. Ernst and Banks 2002). Each additional dimension should be expected to make an exponential contribution to the problem size.

Individuating Inferences

One thing we haven't discussed yet is how to individuate inference problems. As it turns out, this question matters a great deal. This is because inference problems are *much* more than the sum of their parts. So far we've been assuming that if perception represents the dimensions of color, lighting condition, shape, and location, then it must recover these in a single inference problem. But recovering them in a set of smaller inference problems is exponentially less costly.[34] Imagine, for example, that perception were to solve two inference problems, one to recover the color and lighting condition of an object and another to recover its shape and location. Using the same figures we used above, but recovering the surface color and lighting color for three objects and, separately, three object's shapes and locations, would deliver a hypothesis space of approximately

[32] Here I am calculating (3/100)^36 – or 12 dimensions over 3 objects. This assumes that shape representations are parameterized along 3 natural, continuous dimensions. This is almost certainly not the case. The actual dimensionality will be higher, and the resulting proportion of acceptable hypotheses will be smaller. We are considering 'astronomical' anything > 10^30, see Section IV.

[33] Which is not just successive location (see the waterfall illusion).

[34] Based on our assumptions so far.

10^51 hypotheses; about a million trillion times fewer than the 10^90 that results if we jointly solve for all of these dimensions.[35] These savings only get more dramatic as the overall dimensionality grows.

One might wonder whether perception could limit the costs of inference by adopting a divide and conquer strategy of this kind, in effect holding that perception is composed of many distinct modules responsible for each of the different sets of properties discussed above. This broad outlook on vision was made famous by Marr's foundational work on vision (Marr 1982) and has many contemporary adherents.

The problem with such an architecture is what is lost when larger inference problems are broken up into smaller problems in this way. In such cases, the sum of inference problems is no longer sensitive to the *dependencies* between the dimensions housed in separate problems (more on this in a moment). The loss of sensitivity to these dependencies matters because human-level performance requires this sensitivity (otherwise, color, shape, and location cannot be accurately recovered), and, unsurprisingly, human vision empirically exhibits it (as will become clear shortly). The rest of this subsection will spell out this reasoning more carefully.

Dependencies, as may be clear from the above, are the relationships between dimensions such that information about one dimension bears on the probable values of another. Sensitivity to dependencies is necessary if inference is to arrive at an internally consistent percept. For example, if one large object stands between another object and a scene's source of illumination, then the second object is likely to be cast in the first's shadow. This in turn influences how the intensity (and spectral profile) of the light reflected off the second object is interpreted, as object color or lighting condition. Conversely, if the light bouncing off an object of unknown location is reflecting light that is darker than expected, this could be evidence that the object is in shadow, providing information about its location. Such dependencies between dimensions (in this case location and lighting condition) are invisible when inference problems are broken up into their component parts. In such cases, assignments of probable values of color must be made independently of assignments about location, leading to inconsistency.

If such inconsistency is kept to modest levels, it might be a reasonable price to pay for tractable inference, but it does not seem to be the strategy that human perception takes. This is because perception is, in fact, sensitive to a great many dependencies between perceptible dimensions, including dependencies between all of the dimensions used in the toy example above. Sensitivity to these particular dependencies can be seen through a series of established psychophysical results. (Such results will naturally not show that perception is

[35] For color and lighting color, that's ((100)^6)^3 = (10^11)^3 = 10^33. For shape and location, ((1000)^6)^3 = (10^17)^3 = 10^51. 10^33 + 10^51 ≈ 10^51.

sensitive to the dependencies between all of the dimensions it represents, but will show that a divide and conquer strategy is insufficient for tractability, as the dependencies which are represented are sufficient to establish astronomical costs given our other assumptions.)

Start with color constancy – Objects in the world are seen as having a stable color, despite changes in lighting condition between indoors and out, across changes in weather and time of day. This fact is quite surprising when one considers just how much the light hitting your eye differs under these conditions. A lump of coal in bright sunlight reflects about as much light as white chalk indoors, but the chalk appears bright white and the coal jet black. This phenomenon, known as color constancy, is accomplished by jointly inferring color and lighting condition so as to find a consistent assignment of values to those dimensions (Tokunaga & Logvinenko 2010). If a lot of light is hitting the retina, for example, this could be because the object reflects most light (as in the case of chalk) or because it is intensely lit (as in the case of coal in bright outdoor light). By doing joint inference over these dimensions, perception can ensure that it is not double counting the properties of the proximal stimulus – which might result in seeing the coal outside as bright white. Color constancy then, is perceptual sensitivity to the dependency between color and lighting condition.[36]

Just as with color and lighting condition, all four of the dimensions we've discussed so far are jointly confounded in the retinal stimulus and so conditionally dependent on one another. For example, different shapes in different lighting conditions give rise to different patterns of coloration across an object. If information about probable lighting sources is present, either from a prior or from further cues in a scene, then the pattern of coloration can be used to infer the object's shape. In a phenomenon known as 'Shape from Shading,' the visual system does just this. A classic study showed participants 2D shaded circles, either darker on the bottom and lighter on top or vice versa. Participants saw the light-on-top circles as convex 3D reliefs while seeing the dark-on-top circles as concave recesses, demonstrating both a visual prior that light comes from above and a sensitivity to the dependency between lighting condition and shape (Ramachandran 1988, see Figure 1 for illustration). Sensitivity to the dependencies between color, lighting condition, and shape extends to the location dimension and to the properties of other objects as well. When multiple shaded objects provide further cues to lighting direction, participants can be induced to assign different locations to an unobserved lighting source (Morgenstern et al. 2011). Similarly, scenes with cues suggestive of multiple lighting sources induce global percepts of objects with shape properties consistent with those lighting sources (Wilder et al. 2019).

[36] Really, the conditional dependency between color and lighting condition, conditional on a given retinal input.

Collectively, these effects illustrate perceptual sensitivity to the dependencies that exist between color, lighting condition, shape, and location. Insofar as color and lighting condition are jointly dependent on one another (by color constancy), lighting condition is dependent on shape and location (by light source, shadow, and mutual illumination), and the locations of objects and light sources are dependent on the shape and color of objects (by the flexibility of the illuminant prior), there cannot be any consistent independent recovery of these attributes. Rather, they must be recovered jointly.

A vivid illustration of this joint inference can be found in the bistable chromatic Mach card (Bloj et al. 1999, Harding et al. 2012). In this effect, a folded card with two colored sides is shown to participants. One side is painted white and the other magenta. The card is folded in a concave fashion, with the edges of the paper protruding, and presented to the viewer head-on. Viewed at this angle, the card can be seen as either concave or convex. Because the card is actually concave, the two sides mutually illuminate, with light from the magenta side casting a pink glow on the white side. When participants see the card as concave, all of this is perceived veridically – the card looks concave, the sides white and magenta, and the white side cast in pinkish light. When participants see the card as convex however, one side is perceived as magenta and the other side as light pink (i.e. having a light pink surface color). In this case, the pinkish coloring that vision had originally attributed to mutual illumination between two facing sides is now seen as the much darker surface color of a second painted side. The bistability of the chromatic Mach card vividly illustrates human visual sensitivity to the dependencies between color, lighting condition, shape, and location (mediated by mutual illumination).

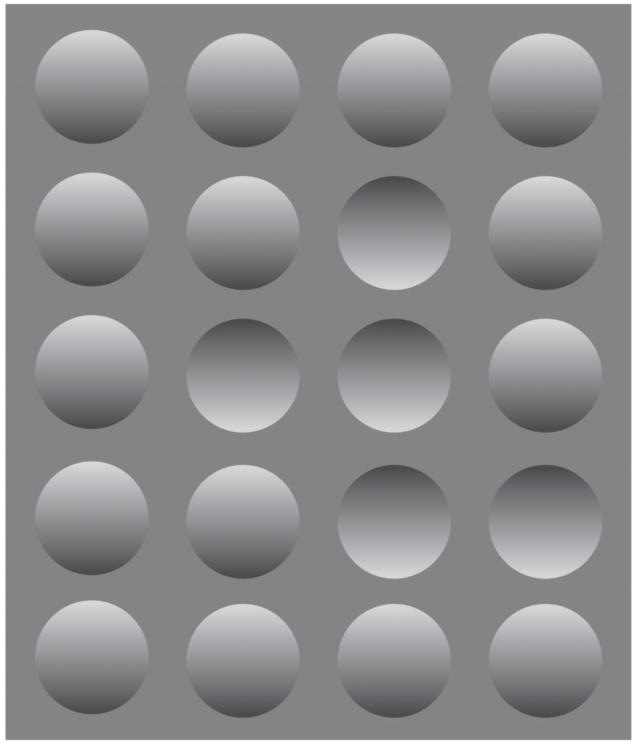

Figure 1: Typical shape from shading stimuli – Shape (either convex or concave) is assigned to multiple objects under the assumption of a single illuminant governing them all. This assumption is defeasible, as discussed in (Morgenstern et al. 2011, Wilder et al. 2019).

Sensitivity to these dependencies shows that human perception cannot be using a simple divide and conquer strategy to head off exponential costs. But what about a modular strategy followed by a recombination stage? There are lots of ways that such a strategy could work, but they all fall into two broad categories – independent computation of dimensions followed by *principled* combination of those values into a coherent hypothesis, and independent computation followed by *heuristic* combination. We'll look at each of these in turn.

Take the first case, of independent recovery followed by principled recombination. When this strategy is deployed, the problem is first broken up into small subsets of dimensions which are jointly inferred, with exponential savings for breaking up the larger inference problem. The outputs of these sub-inferences are then recombined into a full hypothesis in some principled fashion, such that the end result is the same as if inference had originally been computed over the full set of dimensions. Illustrative examples of this approach come from the literature on 'Bayesian cue combination.' In a typical Bayesian cue combination study a model is proposed on which independent measurements of some perceptual dimension are combined in a way that is sensitive to the uncertainties in each measurement. These independent measurements are then combined analytically, often by multiplying gaussians. In one famous study, due to Ernst and Banks (2002), subjects were asked to assess the height of an object presented to them both visually and haptically. This was accomplished by allowing subjects to simultaneously touch an object with their hands while viewing it through a window of varying opacity, blurring the image of the object beyond. The authors showed that subjects' judgements of size reflected information from both vision and touch. Intriguingly, subjects' final judgements were also sensitive to the uncertainty in each of the input modalities, with the more certain (lower variance) channel having a greater 'weight' in the final judgment. Vision was relied on more by default, but subjects' judgements reflected greater weight placed on haptic information as visual inputs were made noisier (by increasing opacity of the viewing window). Finally, the uncertainty of subjects' final judgements was always less than the uncertainty of the measurement from the more reliable modality, suggesting that information from both channels was in fact being integrated, rather than information from the less reliable modality being thrown away.

What's interesting about this work for our purposes is the way in which information is integrated. In these models, inference (the process of considering and evaluating hypotheses discussed above) is entirely eschewed. Rather, information is combined analytically – in this case by multiplying two normal distributions representing independent visual and haptic measurements of the relevant value.[37] When measurements are combined analytically in this way, the full costs of inference are avoided, leaving only the costs of inference over the subsets of dimensions combined together and the trivial cost of multiplying gaussians.

Despite the promising start, approaches of this kind face several problems that severely limit their generality, and hence their viability as models of perceptual inference.[38] Here I'll focus on just one such problem. In cases of Bayesian cue integration, an analytic solution to integrating the outputs of partial inferences is available only when integration is mandatory. So, in the case of Ernst and Banks above, subjects' perceptual systems were able to recognize that the haptic and visual input came from the same object, and so it made sense to integrate information from both senses. But we often find ourselves touching and viewing distinct objects,

[37] Indeed, Bayesian cue combination is typically framed as independent measurements of a single dimension, rather than inference over multiple dimensions. What it mimics is true inference over low level haptic and low level visual dimensions in order to recover the height of an object. Cue combination and inference output the same value if the relevant uncertainties over size actually are independent and gaussian in the full model.

[38] It's unclear if the authors of studies of this kind ever intend their models to be understood in this realist way, as models of the actual computational processes by which the brian solves these inference problems, rather than as demonstrations of the optimal use of information by whatever process the brain actually implements (that is, whether the models are ever intended as Marr algorithmic level models). The concerns I'll offer here give us reason to doubt that the brain actually computes inference in the way described by these models, but not to doubt that the brain is sensitive to dependencies in the ways the model describes.

and in these cases we do not integrate information from haptic and visual channels (Kording et al. 2007). The question then is, how does perception know which case it is in (whether the objects are distinct or identical) and therefore whether it should integrate? Models of this integration-decision require nothing less than full inference over the relevant dimensions to determine whether a single cause of haptic and visual inputs, or distinct causes, is more likely (see Kording et al. 2007, Beierhold et al. 2007). In this case, the exponential costs of inference cannot be avoided by analytic integration.

The challenge for the approach above is that delivering the outputs of inference in the general case seems to require inference. A natural thought at this point is that there might be some heuristic method for integrating disparate sub-inferences – here a heuristic method is defined as one that integrates sub-inferences well enough to meet the needs of human vision, but is not guaranteed to work in all cases. Delivering such a heuristic is easier said than done. To get a sense for the difficulty, consider what heuristic means of integration would give rise to behavior exemplified by the Mach Card described above. What general heuristic tells us when colored light should be seen as part of the lighting condition, rather than object color? Or could tell the visual system how to update its assessment of an object's shape as a function of those assignments? Or recover the number and location of lighting sources based on the shadows cast on objects of disparate shape? The sheer number of ways that the dimensions of shape, color, location, and lighting condition might depend on one another makes the prospect of a heuristic method of integration adequate to human vision itself an exponentially vanishing prospect. At a minimum here, we can note that no such general heuristic method of integration has been proposed in the litterature.

Our assessment of the viability of these proposals is, of course, subject to change. Perhaps a heuristic approach to the problem of inferential integration will come along, and one should be taken seriously if and when it does appear. For the moment, however, there does not seem to be an alternative to doing inference, which minimally must respect the dependencies described in our toy inference problem and illustrated by the bistability of the Mach Card. If this is right, then the assumptions we've explored so far are sufficient to land the costs of perceptual inference in astronomical territory. Any would-be explanation of the tractability of perception must then account for how those assumptions can be challenged, allowing such costs to be avoided. In the following section, we look at what it would take to provide an explanation of the tractability of perception along these lines.

VII. How (And How Not) To Explain Tractability

How to Explain Tractability

It would seem then that we've reached an impasse. By our lights the intrinsic costs of perceptual inference scale exponentially in dimensionality and a mere subset of the dimensions involved in perceptual inference run those costs into astronomical territory. For perception to be tractable, however, the costs of performing inference must not be astronomical. At this point we need to stop and take stock of the assumptions that got us here and ask ourselves if any of them can reasonably be denied.

As a reminder, these assumptions were threefold: (1) that good hypotheses are found, relative to a reasonable performance criterion. (2) that the costs of evaluating hypotheses are relatively fixed. And (3), that nothing is known about the distribution of good hypotheses in the hypothesis space. We discussed (1) and (2) at length in Section V.[39] That leaves (3). For (3) to be false would mean that perception has information, in advance of inference, about the distribution of plausible hypotheses in the hypothesis space. If perception *were* to have such prior information, this information could be used as a guide when exploring the hypothesis space. While drawing hypotheses randomly entails exponential growth in the expected number of hypotheses sampled before finding a good one, guided exploration of the space does not – in the guided case, the costs would depend straightforwardly on the quality of the information used as a guide.[40] To deliver tractability, this information must be good enough to find criterion-meeting hypotheses from among astronomical numbers of options in fewer than K steps.

Take 'sampling' to describe the choice that any inference algorithm must make as to where to look for good hypotheses in the hypothesis space.[41] We can call the outcomes of these decisions an algorithm's 'sampling dispositions.' When these dispositions are informed by information about the distribution of good hypotheses in the space, we'll call them *intelligent sampling dispositions* (ISDs). With this concept in hand, we can now offer a precise statement of the problem of the tractability of inference:

---

[39] P. 30ff

[40] See Chatterjee & Diaconis (2018).

[41] In this case, we're using the term to describe something broader than sampling in the technical sense that is relevant to the Monte Carlo inference methods that might be familiar to some readers. Sampling in our sense includes any way that an inference algorithm might go about selecting promising portions of the hypothesis space, including those in non-Monte-Carlo inference algorithms, such as variational methods.

The Challenge of Tractable Inference: The challenge of explaining how perceptual inference is tractable by accounting for the intelligent sampling dispositions at work in perceptual processing.

For the rest of this paper, I will defend the claim that the challenge of explaining the tractability of perception is the challenge of explaining how perception comes to have intelligent sampling dispositions (ISDs) sufficient to avoid astronomical costs when performing inference in an astronomical space of options.[42] While candidate ISDs abound, delivering on such an explanation that is up to the task of perceptual inference is far easier said than done.[43] What makes it difficult is that the location of plausible hypotheses is not fixed, but is rather sensitive to the specifics of the problem instance at hand. We see very different scenes in the course of our lives, and which scene we're looking at on any particular occasion dictates where the plausible hypotheses are to be found.

To see why delivering such intelligent sampling dispositions is difficult, it is helpful to see why one popular idea, that the perceptual system embodies 'natural constraints' on perceptual scenes, is not a solution.[44] The idea of natural constraints is the idea that the perceptual system has access to (implicitly or explicitly represented) information about how the world typically is. The canonical example here is the visual system's sensitivity to the fact that light typically comes from above (see discussion of shape from shading in Section VI above). That the visual system possesses such a prior may be true as far as it goes. But such a prior, even if it's used to inform sampling, is unlikely to address the issues of computational tractability discussed here. This is for the simple reason that human vision in fact recovers any number of different lighting sources and lighting directions, and recognizes fine-grained local differences in lighting condition, such as shadow and mutual illumination (all while respecting the dependencies between these dimensions and many others, see Section VI). The mere starting assumption that lighting is singular and comes from above does not save vision from the requirement to be sensitive to a vast number of other ways that lighting could be, including more fine-grained

[42] See Schulz (2012) for a similar thought in the case of cognitive inference in children.

[43] Any inference algorithm that delivers a speed advantage over exhaustive search or uniform sampling will have some ISDs that are responsible for its speedup. This includes algorithms making use of the idea that good hypotheses tend to be near one another (e.g. local MCMC), that good assignments of values to variables will be high probability conditional on good assignments to other variables (e.g. Gibbs sampling), that the posterior landscape is smooth (e.g. Hamiltonian MCMC, Variational Methods), etc.

[44] Or at least not a solution on its own. Note that many of the proponents of information encapsulation are also proponents of natural constraints (the information in perception has to come from somewhere, after all) and so already accept that perception has prior information about its domain. I expect for this reason that many will be broadly sympathetic to the idea that more information is present in the form of ISDs.

ways consistent with light coming from above, and it is this requirement that entails astronomical computational costs.

While natural constraints are not themselves enough to deliver computational tractability, they are the right kind of thing. That is, they are sources of information, prior to inference, about which perceptual hypotheses are likely to be good. What's needed to account for tractability is much stronger sources of this kind of information. In contrast to natural constraints, which embody information about which hypotheses are plausible *in general*, what is needed for tractability is more fine-grained information about the distribution of plausible hypotheses for the problem instance at hand.[45]

Why Information Encapsulation Does Not Explain Tractability

Now that we better understand the sources of intractability in perceptual processing and what is needed to avoid astronomical computational costs, we're also better able to see why information encapsulation is not an explanation of tractability. The main idea here is that the costs intrinsic to perceptual processing are vastly larger than those associated with information access, and this difference in size undermines any intimate explanatory connection between encapsulation and perceptual tractability. This is the main idea, but I don't expect the reader to be convinced just yet. As always, the devil is in the details. In what follows, we'll go through a series of things it might mean for information encapsulation to explain tractability – including the possibility that information encapsulation is sufficient for tractability, that it is necessary, or that it is a difference maker. We'll see how our new appreciation of the challenge of accounting for perceptual tractability allows us to definitively rule out versions of the EET on which encapsulation is necessary or sufficient for tractability, while leaving us with strong reasons to be skeptical that it might be difference maker.

Start with sufficiency. Could avoiding the costs of information access by way of encapsulation be *sufficient* for the computational tractability of perception? Based on what we've said so far, the answer to this is clearly no. This is because ISDs are necessary for computational tractability, and a perceptual system could be encapsulated from a cognitive system without also possessing ISDs. For example, a simple model aimed at doing the inference described in Section VI might receive no inputs from any external computational system (and so be encapsulated) and yet lack any ISDs. In the simplest case, it could perform inference by sampling randomly from the space of possibilities. Such a model would be encapsulated, but inference in it would be straightforwardly

[45] That is, not merely a good prior, but a good estimate of the posterior. For the recurring distinction between dimensions and information, see p. 38.

intractable, running up against the astronomical costs of inference. So encapsulation is clearly not sufficient for computational tractability.

How about necessity? Could information encapsulation be necessary for computational tractability? Here too I think the answer is no, but before arguing for this, it's worth first seeing why this idea commands so much appeal. There is a ton of information in cognition, from random facts about people, such as names and political persuasions, to the habitats of animals, to memories of your grandmother's garden. Perception, for its part, has to operate very fast, on the order of tens or hundreds of milliseconds, as we saw before. What's more, some kinds of very demanding search are certainly intractable. Take, for example, what we might call 'full relevance search.' By full relevance search, I mean sorting a list of information into those entries that are relevant to an inference problem and those that are not. In the limit, this requires performing the full inference problem once with each subset of the entries on the list and comparing the results to see which entries make a difference (in different combinations) to the outcome of the inference, in order to determine which entries are relevant to the task at hand. Such an operation is likely to scale super-exponentially, since it involves inference (which scales exponentially with dimensionality) being performed as a subroutine exponentially many times (as a function of the size of the list). Search of this kind would of course be intractable. If perception were required to do an exhaustive relevance search through cognition in the course of each perceptual inference, then there can be little question that it would be intractable.[46]

This is all true as far as it goes. But it is also very far from establishing that encapsulation is necessary for tractability. This is for three reasons. First, search does not have to be exhaustive, going through the entire mind to guarantee that it has returned all of the relevant information, in order to violate encapsulation. Search methods might search some portion of the database that merely sometimes has relevant information (say, 'search only memories from the past 24 hours'), or might search in a way that could access the entire database, but with a limited amount of time in which to do so ('search everything but stop after 100 milliseconds'). Other ways of limiting search exist as well. Instead of circumscribing search on the basis of the store or the duration of the search process, search could be limited by properties of the information being accessed, say returning values based on their place in the full list of entries (even very simple organizations of lists keep search costs sub-linear)

[46] Full relevance search of this kind seems to be what Fodor (1983) has in mind when he writes, "the point of the informational encapsulation of input processes is not—or not solely—to reduce the memory space that must be searched to find information that is perceptually relevant. The primary point is to restrict the number of confirmation relations that need to be estimated as to make perceptual identifications fast" (p.71). See Fodor (2000) for similar arguments about the intractability of relevance search.

or on the basis of their syntactic features (say, ‘return only those memories that explicitly encode the color of this object’). If perception does in fact search through cognition, it could limit its search in any of these ways, making the costs of exhaustive relevance search irrelevant.[47]

Second, not all kinds of search that return some relevant information require sorting that information into relevant and irrelevant entries, and it is often better not to do so. Consider an over eager search strategy that returns some relevant information and much that is irrelevant. If we do inference with this information, the outcome is the same as if we’d done inference without the irrelevant information (that’s what it is for the information to be irrelevant!). As for computational costs, the costs are no more than if we’d first sorted the list into relevant and irrelevant entries and accessed only the relevant ones (since the information has to be accessed in both cases – either to be fed directly into inference or to be sorted) and are often much less (since the super-exponential costs of sorting are neatly avoided in the over eager case). Inference itself is not more expensive with the irrelevant information, since the costs there are dictated by dimensionality, not information.[48] So search for relevant information need not be the super-exponentially scaling relevance search of the kind envisaged above.

Finally, search strategies that reliably return relevant information without exhaustive relevance search are not an idle theoretical possibility. Rather, search strategies of this kind are a fixture of the modern era, making searching through even extraordinarily large databases fast and efficient. A typical Google search, for example, searches Google’s copy of the internet, an enormous body of information, and returns general relevant results at an average latency of 500 milliseconds.

With all of this in mind, we can now see why encapsulation cannot be necessary for perceptual tractability. Consider a perceptual system with the following property: after coming up with an initial guess as to the identity of an object, it runs a Google search to find the typical color of that object, and uses this as an additional input into color and identity processing. This system would be unencapsulated, in the vein of anti-encapsulation interpretations of color processing effects in people (MacPherson 2012). More importantly for our purposes, if inference in this system was tractable before adding the search, then it will be tractable

[47] Note that in Section III we assumed for the sake of argument that information access could be tied to the costs of search, despite the possibility of search without access (say, if cognition does search and sends information to perception as an expectation prior to inference, see Kok et al. 2012). I am not reneging on this deal – the costs of search are still at issue – but rather pointing out that search does not entail exhaustive search.
[48] See Section V, p. 37 for the distinction between information and dimensions.

afterward. The possibility of such a case shows that, at least based on our current evidence, encapsulation can't be necessary for tractability.

Here I want to be clear about what I am saying and what I am not. The point is not that search in the mind might work just like Google search – very likely this is an unrealistic model. The point is rather that Google search gives us a proof of concept that some searches over very large databases are nevertheless very cheap. In a few short decades of computer science, human ingenuity has already hit upon cheap ways of doing large scale search. In light of that, we would need a very strong argument to convince us that cheap ways of doing search were out of reach for evolution. And without such an argument, we should not believe that avoiding the costs of search is necessary for tractability (cp. Clark 2002 for a similar point).

If encapsulation is neither necessary nor sufficient for tractability, then in order for the EET to be true encapsulation must at least be a difference maker. To be a difference maker it must be the case that, given all the facts on the ground, if perception were unencapsulated, then it would be intractable.[49] Here the thought would be that engaging in information access is a discretionary line item in the brain's computational budget for perception, and one that pushes perception over budget after all the essential line items are paid for. The question then is, what reasons could we have for believing that information access is such a decisive line item? These reasons break down into two categories. First, we could have reason to believe that those costs are a large part of the final budget for tractability, once all the strategies that evolution has employed to keep costs low in search, inference, etc. have been taken into account. (This would mean that the costs of information access would also be a big part of the final budget of K once a much tighter bound had been set on K). Second, even if the costs of information access are not a big part of the budget, we might nevertheless have reason to believe that they are a small, but critical part of the budget – the final line item that just tips the balance and pushes us over budget. Here, as above, I'll argue that the vast difference in scale between the problems of inference and access undermines either case for believing that avoiding the costs of access will be a difference maker.

Take the first possibility. Do we have any reason to believe that the costs of access will be a large part of the final budget, once we've figured out all of the optimizations evolution has employed to keep the costs of both access and inference down? In evaluating this admittedly very challenging question, start with a sociological fact: When computer scientists evaluate the complexity of a program which is the sum of a non-exponential term and an exponential term, they tend to ignore the non-exponential term. They don't do this out of a sense of wanton

[49] That is, for a natural analysis of what it is to be a difference maker, which is to be a necessary condition holding everything else about the system fixed. See p. 21 for a brief discussion of what it is to be a difference maker and how this differs from both necessity and sufficiency.

violence toward an accurate representation of the complexity of the program, but rather because, empirically, when computational costs are the sum of a non-exponential and an exponential or greater term, the contribution from the non-exponential term tends to be negligible. That is, if the costs of running a program are $n^x + 5x$, this is typically very well approximated by n^x.[50]

New insights into how full inference and exhaustive search might be approximated could, of course, change this. If evolution has been tremendously successful at keeping the would-be exponential costs of inference low, while finding few or no strategies to lower the would-be linear costs of search, then this could change the relative proportions of the budget that go to each term, leaving search and access the larger chunk of the final budget. There is no doubt that this is possible. But it does not seem particularly likely. For one, the starting costs are so different that the successes in lowering costs would have to be remarkably one-sided. For another, as we saw just above, the current state of affairs paints just the opposite picture – we currently know of many methods for making theoretically cheap search even cheaper, while we have very few insights into how theoretically expensive inference could be made much less expensive. At the very least then, we have no positive reason to believe that encapsulation will turn out to be a difference maker for perceptual tractability by way of being a large portion of the final budget for perceptual processing.

If the costs of information access are not a large part of the budget, we could still have reason to believe they are a difference maker if we have reason to believe that they are a small but critical part of the budget – the final expense that pushes us over budget once all essential operating costs have been paid. In evaluating this possibility, consider one final time the difference in size between our exponential and our linear terms in the theoretical costs of unencapsulated perceptual processing. If our thinking so far in this paper is on the right track at all, then the final costs of information access are likely to be a drop in the bucket compared to the final costs of inference. The theoretical result (based on the difference in theoretical costs) is clear cut, and the convergent empirical evidence (based on the current ease of search, discussed above, and current difficulty of inference, briefly surveyed in Section II) is at least suggestive of a significant difference in the computational costs associated with these two problems. Given our current evidence then, believing that the costs of information access are a small but still critical portion of perception's computational budget would require believing that

[50] The exception to this is when x is small, in which case the linear term will dominate. That does not arise in this case, since the variables are distinct (the exponential term is exponential in dimensionality, while the linear term is linear in the number of entries that must be searched) and the exponential term is fixed empirically by the dimensionality of the inference problem (see Section VI).

these costs are going to be the proverbial drop that makes the bucket overflow. This is certainly possible, but we have no positive reason to believe it!

We've seen that the difference in size between the problem of inference and the problem of access rule out certain versions of the EET (that avoiding the costs of access is necessary or that it is sufficient) and give us considerable reason for skepticism about others (that avoiding the costs of access is a difference maker). Thus we've seen that the original motivation for the EET, what we earlier called the Haystack Idea, can be safely laid to rest. There is, however, one final difference making role for encapsulation that should be examined. This is the possibility that encapsulation might be a difference maker, not by allowing perception to avoid costly search, but rather by being a crucial part of the ISDs which are themselves critical to the tractability of perceptual inference. In this case, proponents of encapsulation would acknowledge that the costs of search are likely insignificant in explaining tractability, but would turn to seeing encapsulation as a plausible contributor to limiting the actual costs of inference. I'll briefly argue that even this revitalized version of the EET lacks sufficient motivation.

As a way of easing into it, start with a simpler thesis, that information encapsulation is necessary for ISDs. Based on what we've established so far in this paper, we know that this can't be the case. This is because having an amazing perceptual prior, one with strong expectations about what you're likely to see when,[51] is sufficient for ISDs. And it's possible to have an amazing prior while not being encapsulated from cognition. If a perceptual system came equipped with such a prior (by way of evolution or perceptual learning), but were unencapsulated (accessing select information from cognition, say, just color memories), then this system would exhibit strong ISDs despite being unencapsulated. At least based on what we know right now then, encapsulation can't be necessary for ISDs.[52]

Now consider the possibility that encapsulation might be a difference maker – making some critical contribution to perceptual ISDs holding all other facts about the system fixed. To evaluate whether this is plausible, start with how it is that ISDs provide for tractability. They do so by helping to locate plausible hypotheses from among an astronomically large expanse of random possibilities. At first pass then, any information that might help the system locate plausible hypotheses is likely to result in computational savings. In the next section, I'll argue on these grounds that there are likely to be many cases where information from cognition could significantly *reduce* the costs of perceptual processing. (The gist of those examples is that cognition can sometimes propose reasonable solutions to perceptual inference problems, in virtue of sometimes

[51] And therefore a good approximate posterior.

[52] There is also no question that it's not sufficient for ISDs, since it's categorically the wrong kind of thing to deliver ISDs – not a source of information, but merely a prohibition on one.

possessing veridical information about what we are likely to be seeing.) For the moment, however, let's limit our attention to what it would take for information from cognition to significantly increase the costs of perceptual processing by way of influencing ISDs. To do so, cognition would have to contribute information that is not just sometimes wrong about what you are seeing, but rather systematically and relentlessly misleading about what you might be seeing.

To wrap our minds around this point, start with a related downside to cognitive influence on perception that has been discussed in this literature. Some authors have argued that human perception is susceptible to cases of 'wishful seeing' in which someone's idea of what they would like to see influences their perceptual interpretation of ambiguous stimuli. So, if I am looking for the mustard in my fridge, I might briefly misperceive a lemon in the fridge as mustard (Siegel 2017). Wishful seeing, if it happens, influences the outcome of perceptual inference – the lemon looks to me (perhaps briefly) as if it were mustard. When the outcome of perceptual inference is less accurate then it would have otherwise been, wishful seeing has an obvious *epistemic* cost. What we're imagining here is slightly different. We're imagining a case where cognitive influences have a meaningful *computational* cost.[53] What we should have in mind then is a case like wishful seeing but which holds fixed the final outcome of perceptual processing. In this case, cognition would first offer the mustard hypothesis, and perception would check it against the data, perhaps rejecting it because the mustard hypothesis sits poorly with the absence of any noticeable label or cap on the yellowish figure. Finally, perception settles on the lemon hypothesis, as it would have if there had been no effect of cognition.

Here, even though perception avoids any epistemic cost by ultimately settling on the same output hypothesis, the proposal of the mustard hypothesis creates an unnecessary computational cost – that of evaluating and rejecting the false hypothesis. It stands to reason then that, if there were a sufficient number of such unnecessary hypotheses proposed by cognition, they could run up the costs of perceptual inference. Encapsulating perception from cognition's misleading proposals could then be an important part of how the ISDs deliver tractability.

The problem with this line of reasoning is that, if the implausible hypotheses proposed by cognition are just one or a few, then the costs they impose are negligible relative to the size of the perceptual inference

[53] Of course, both epistemic and computational costs are relevant to tractability, since tractability is relative to both a budget, K, and a performance criterion (see Section III). If cognitive penetration would impair performance to a sufficient degree by way of effects like wishful seeing, then technically such penetration could make perception intractable by way of decreasing performance beneath human levels, rather than increasing costs above human levels. I take it that the core idea behind the EET as it has been defended is that cognitive influences would make perception more expensive, not less accurate, so I won't consider this back route to the thesis here.

problem. Since we should believe that, even with great default ISDs, perception is likely to have to evaluate many hypotheses, the addition of a handful from cognition seems unlikely to be difference making. And if the hypotheses are not implausible, then evaluating them may contribute to veridical perceptual inference. For cognitive influences to pose a threat to tractability by this route then, the proposals from cognition must be both very numerous *and* systematically misleading. While it may be easy to imagine that cognition, were it allowed to, might occasionally send perception an implausible proposal, the idea that it might be a source of implausible proposals on the scale needed to threaten perceptual tractability, where the default requirement is to navigate an exponentially large hypothesis space, is a heavy lift. Here again, at least absent some positive argument in its favor, we should not believe that this is the case.

I'll say more about the relationship between encapsulation and ISDs in the following section. At this point however, we should take stock of where we've gotten. We've seen that encapsulation is neither necessary nor sufficient for computational tractability. It is not sufficient because ISDs are necessary and a system can be encapsulated without exhibiting any ISDs. And it is not necessary because the kinds of information that allow for tractability can exist in perception even if it is unencapsulated. We've also seen that the costs of information encapsulation are unlikely to even be a difference maker to perceptual tractability due to the massive difference in size between the costs of information access and the costs of inference itself. After establishing that the costs of information access are unlikely to motivate encapsulation, we finally asked if encapsulation might be critical to tractability by way of being critical to ISDs. And we found that defending such a view requires believing that the proposals influenced by cognition are not merely sometimes false, but both very numerous and systematically implausible. The idea then that tractability considerations motivate encapsulation should be laid to rest. Many questions remain, however. In the next section, I use some of the tools we've laid out in this paper to explore the future of tractability arguments.

VIII. The Future of Tractability Arguments

The Dimensionality Restriction Hypothesis

At the end of the day, we don't know how perception is tractable, and this limits what we can say with confidence about the bearing that various cognitive effects might have on tractability. But we are not totally in the dark either. For example, we have an understanding of the sources of computational costs and the factors that influence them. Any cognitive effect that threatens to make those prima facie costs exponentially worse should be regarded with suspicion. Similarly, we know the form that a solution to intractability must take – it must offer a theory of the intelligent sampling dispositions that allow perception to navigate its vast hypothesis space. Such dispositions are built on information. Any potential source of this kind of information, up to and including cognition, should be regarded as potentially part of the solution.

Start with an example of the first kind. The unfortunate truth of computational costs is that, while it can be difficult or even impossible to make a problem easier, it is always possible to make it harder. Certain kinds of cognitive effects could make perception's problem much harder. Take for example the 'enrichment' of perception by cognition. Some authors have suggested that cognition might enrich perception, in the sense of expanding perception's representational capacity to include dimensions previously represented only in cognition, for example in the process of developing expert perception (Siegel 2010). Others argue on empirical grounds that perception is dimensionality restricted, operating over an (at least synchronously) limited set of dimensions, in contrast with cognition, which is dimensionality non-restricted (Green 2020). The ideas laid out in this paper suggest that there may be more a priori considerations relevant to this debate as well. Since the prima facie costs of inference scale exponentially with dimensionality, adding dimensions, whether from cognition, perceptual learning, or by any other mechanism, could dramatically increase the costs of perception. Such an effect could form the basis for a tractability argument against cognitive enrichment effects and in favor of the dimension restriction hypothesis. Making this case rigorously would require careful treatment, but the possibility of such an argument follows naturally from the framework developed here.

Veridical Information From Cognition

Another species of future tractability arguments could put information encapsulation on the defensive. As we saw above, what's needed to account for the tractability of perception are sources of information which help steer perceptual inference toward promising hypotheses. A natural question to ask, then, is whether information *from cognition* could support tractability.

Consider the following case. You are wandering around in a jungle and see an ambiguous form in the branches. Naturally, there are virtually countless possible visual interpretations of the visual scene. Now imagine that you know that you're in panther territory. This key bit of information from cognition could be used to guide sampling, allowing vision to arrive at an interpretation of the visual scene much more quickly. A tip of this kind could easily be the difference between visually detecting and missing something that was really there.

How, exactly, could the abstract belief that one is in panther territory be used to guide visual inference? The technical proposals are too much to get into here. But the underlying process is not that different from what you would naturally do if I were to ask you to close your eyes and imagine a panther in that tree. Then to reset and imagine another, distinct scene meeting the same constraint. And then another. In each of these cases, you are sampling from a space of possible scenes, under the constraint that they feature a panther in the tree. This cognitively constrained distribution is far more peaked than an unconstrained prior distribution over all possible scenes, with or without panthers, thereby guiding visual inference toward the hypotheses that meet the constraint.[54]

Accounts of roughly this kind have been offered as explanations for the phenomenon of stably resolving ambiguous images (Lupyan 2017, Block 2022). These are images which appear one way at first, say, as an unremarkable brick wall or set of black splotches, but resolve another way when people are given a clue semantically related to the alternative interpretation. Once their more surprising interpretation has been seen, it is often difficult to unsee; a fact which may reflect the visual systems assessment that the new hypothesis offers a better solution to that particular visual inference problem (see Figure 2).

These are not knock down demonstrations of cognition-fed sampling dispositions. Many who have discussed these effects have argued that they are due to attention (Firestone & Scholl 2016, Lupyan 2017, Block 2022 signals openness to this interpretation). Whether attention offers a competing explanation or is merely the mechanism of cognitive penetration is itself an open question (Quilty-Dunn 2019, see Green 2020). Cognitively-driven samples are a computational process while attention is a folk psychological and neuroscientific concept and the relationship between the two is unclear. This is murky territory. My aim in bringing these issues up is not to try to lay them to rest, but merely to illustrate that the tractability considerations that were once taken to require the encapsulation of perception from cognition, may in fact support just the opposite conclusion once the true challenge of perceptual tractability is appreciated. This reversal holds even if, as is likely to be the case, most of the information in the intelligent sampling dispositions is internal to perception.

[54] Note that a procedure like this can work even if the visual system doesn't explicitly represent high-level contents such as 'panther', since the relevant distribution could be a distribution entirely spelled out in terms of low-level properties; those that would trigger recognition of panthers.

Figure 2: Visual inference can be affected by information about what one is looking at. Look at the image aboves and search for anything out of the ordinary before reading this footnote for a hint.[55]

IX. Conclusion

A theory of the architecture of perception must explain how perception is computationally tractable. This paper has argued that information encapsulation, even if true of perception, does not provide such an explanation. This is because of the significantly greater costs of perceptual inference, as compared to information access, which threaten to make the costs of access a negligible proportion of perception's computational budget. After all this, it remains an open empirical question whether perception is encapsulated from cognition, but the encapsulation thesis has lost its computational *raison d'etre*. As a consequence, we should be more willing to accept some of the psychophysical effects reported in the literature as genuine violations of encapsulation. We are at the very least not bound on computational grounds to find ways in which these effects are not genuine effects of cognition on perception. This is, of course, not to advocate for laxity in our psychophysics or analysis, and alternative interpretations of putative cognitive effects should be carefully proposed and ruled out, but it is an argument for a more even prior between encapsulation and cognitive influence as we approach these debates.

The framework for thinking about computational tractability laid out in this paper also has implications beyond the question of encapsulation. For one, we now have an understanding of the sources of computational costs and the factors that influence them. The things that matter to tractability are things like the

[55] On first encounter with this image, most people see an small, bluish rock wedged in a stone wall. Given a hint, such as the quip that 'Sometimes a cigar is *just* a cigar', people see the image differently. (If that is not enough of a hint, try seeing the bluish rock and brown space next to it as a single object, protruding outwards from the wall, with the blue tip farthest from the surrounding rock.)

dimensions, dependencies, and sampling dispositions involved in inference. Information is a resource for limiting computational costs, rather than a liability. With these factors in mind, novel proposals about the architecture of perception can be evaluated for how they are likely to affect tractability. Proposals to the effect that cognition might expand the range of dimensions perception computes over, thereby increasing the dimensionality of perceptual inference problems and threatening to increase the costs of inference exponentially, have an a priori strike against them, while the alternative, dimensionality restriction, has an a priori consideration in its favor. Going forward, we should be more skeptical of, and more careful to explore alternative explanations for, psychological effects which purport to evince such dimensionality non-restriction (in effect, saving for dimensionality non-restriction and other exponentially costly architectural theses the jaundiced eye we have hitherto reserved for purported failures of encapsulation.)

We should also look to develop positive accounts of perceptual tractability. Proponents of information encapsulation were right to think that perception faces a threat of intractability and that reflecting on how such a threat is avoided can be a tool in uncovering the architecture of perception. If anything, this is even more true now; with a vastly larger problem of intractability that is integral to perception's essential function, the demand that an architecture allow for tractable inference becomes a powerful constraint, shaping perceptual architecture throughout.

Deciphering what architectures allow for perceptual tractability is a difficult problem, but we've already made a start – spelling out the general form that such a solution must take. Any account must offer a theory of the intelligent sampling dispositions that allow perception to efficiently navigate the vast hypothesis spaces involved in perceptual inference. Such dispositions are built on veridical information about the distribution of plausible hypotheses throughout the space. In order to deliver human-like perceptual competence, including critically the ability to recover a large number of perceptible properties across a vast diversity of scenes, this information must be opinionated (strongly focusing computational work in narrow regions of the hypothesis space), particular (sensitive to the directly measurable properties of the scene, rather than rigid constraints expected to apply across the board), and veridical (concentrating probability mass around genuinely plausible hypotheses). A theory must tell us where this information comes from and what kind of architecture can gather and deploy it. Any source of information of this kind (up to and including cognition) should be regarded as potentially part of this solution, but the key answers are likely to come from a theory of perceptual learning. Mechanisms of such learning, hitherto treated as something of a black box, are likely to be a critical part of the theory.

References

Alais, David, and David Burr. 2004. "The Ventriloquist Effect Results from Near-Optimal Bimodal Integration." Current Biology 14 (3): 257–62. https://doi.org/10.1016/j.cub.2004.01.029.

Arora, Sanjeev, and Boaz Barak. 2007. Computational Complexity: A Modern Approach. Cambridge University Press.

Battaglia, Peter W., Jessica B. Hamrick, and Joshua B. Tenenbaum. 2013. "Simulation as an Engine of Physical Scene Understanding." Proceedings of the National Academy of Sciences of the United States of America 110 (45): 18327–32. https://doi.org/10.1073/pnas.1306572110.

Bechtel, William, and Adele Abrahamsen. 2002. Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks. Blackwell Publishing.

Beierholm, Ulrik R., Konrad P. Körding, Ladan Shams, and Wei Ji Ma. 2009. "Comparing Bayesian Models for Multisensory Cue Combination without Mandatory Integration." Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference, 1–8.

Block, Ned. 2018. "If Perception Is Probabilistic, Why Does It Not Seem Probabilistic?" Philosophical Transactions of the Royal Society B: Biological Sciences. https://doi.org/10.1098/rstb.2017.0341.

Block, Ned. 2022. The Border between Seeing and Thinking. Oxford University Press.

Bloj, M. G., D. Kersten, and A. C. Hurlbert. 1999. "Perception of Three-Dimensional Shape Influences Colour Perception through Mutual Illumination." Nature 402 (6764): 877–79. https://doi.org/10.1038/47245.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. "Language Models Are Few-Shot Learners." In Advances in Neural Information Processing Systems, edited by H Larochelle, M Ranzato, R Hadsell, M F Balcan, and H Lin, 33:1877–1901. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Carey, Susan. 2009. The Origin of Concepts.

Carruthers, Peter. 2007. The Architecture of the Mind. Oxford University Press.

Chatterjee, Sourav, and Persi Diaconis. 2018. "The Sample Size Required in Importance Sampling." Annals of Applied Probability 28 (2): 1099–1135. https://doi.org/10.1214/17-AAP1326.

Clark, Andy. 2013. "Expecting the World: Perception, Prediction, and the Origins of Human Knowledge." Journal of Philosophy 110 (9): 469–96. https://doi.org/10.5840/jphil2013110913.

Clark, Andy. 2002. "Global Abductive Inference and Authoritative Sources , or , How Search Engines Can Save Cognitive Science." Cognitive Science Quarterly 2 (2): 115–40.

Clark, James J, and Alan L Yuille. 1990. Data Fusion for Sensory Information Processing Systems. Springer Science & Business Media.

David J. Chalmers. 2011. "A Computational Foundation for the Study of Cognition." Journal of Cognitive Science 12 (4): 325–59. https://doi.org/10.17791/jcs.2011.12.4.325.

Davis, Martin, George Logemann, and Donald Loveland. 1962. "A Machine Program for Theorem-Proving." Communications of the ACM 5 (7): 394–97.

Davis, Martin, and Hilary Putnam. 1960. "A Computing Procedure for Quantification Theory." Journal of the ACM (JACM) 7 (3): 201–15. https://doi.org/10.1145/321033.321034.

Descartes, Rene. 2001. Discourse on Method, Optics, Geometry, and Meteorology. Hackett Publishing.

Dodge, Samuel, and Lina Karam. 2017. "A Study and Comparison of Human and Deep Learning Recognition Performance under Visual Distortions." 2017 26th International Conference on Computer Communications and Networks, ICCCN 2017. https://doi.org/10.1109/ICCCN.2017.8038465.

Downey, Rodney G, and Michael R Fellows. 2013. Fundamentals of Parameterized Complexity. Vol. 4. Springer. https://doi.org/10.1007/978-1-4471-5559-1.

Ellis, Kevin, Catherine Wong, Maxwell Nye, Mathias Sable-Meyer, Luc Cary, Lucas Morales, Luke Hewitt, Armando Solar-Lezama, and Joshua B. Tenenbaum. 2020. "DreamCoder: Growing Generalizable, Interpretable Knowledge with Wake-Sleep Bayesian Program Learning," 1–22. http://arxiv.org/abs/2006.08381.

Ernst, Marc O, and Martin S Banks. 2002. "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion." Nature 415 (6870): 429–33.

Feldman, Jacob. 2000. "Minimization of Boolean Complexity in Human Concept Learning." Nature 407 (6804): 630–33. https://doi.org/10.1038/35036586.

Firestone, Chaz. 2020. "Performance vs. Competence in Human–Machine Comparisons." Proceedings of the National Academy of Sciences of the United States of America 117 (43): 26562–71. https://doi.org/10.1073/pnas.1905334117.

Firestone, Chaz, and Brian Scholl. 2017. "Seeing Physics in the Blink of an Eye." Journal of Vision 17 (10): 203. https://doi.org/10.1167/17.10.203.

Firestone, Chaz, and Brian Scholl. 2016. "Seeing Stability: Intuitive Physics Automatically Guides Selective Attention." Journal of Vision 16 (12): 689. https://doi.org/10.1167/16.12.689.

Firestone, Chaz, and Brian J Scholl. 2015. "Cognition Does Not Affect Perception: Evaluating the Evidence for Top-down Effects." Behavioral and Brain Sciences 39 (May). https://doi.org/10.1017/S0140525X15000965.

Flum, Jörg, and Martin Grohe. 2006. Parameterized Complexity Theory. Springer Science & Business Media.

Fodor, Jerry. 1989. "Modules, Frames, Fridgeons, Sleeping Dogs." In Modularity in Knowledge Representation and Natural-Language Understanding. Cambridge: MIT Press.

Fodor, Jerry A. 2000. The Mind Doesn't Work That Way. The Mind Doesn't Work That Way.

Fodor, Jerry A. 1983. The Modularity of Mind. MIT Press.

Geirhos, Robert, Heiko H. Schütt, Carlos R. Medina Temme, Matthias Bethge, Jonas Rauber, and Felix A. Wichmann. 2018. "Generalisation in Humans and Deep Neural Networks." Advances in Neural Information Processing Systems 2018-Decem (NeurIPS 2018): 7538–50.

Gershman, Samuel, Eric Horvitz, and Joshua Tenenbaum. 2015. "Computational Rationality: A Converging Paradigm for Rationality in Minds, Brains, and Machines." Science. papers2://publication/uuid/20A0106C-9CBA-472D-AAFB-69231964766F.

Gigerenzer, Gerd, and Daniel G. Goldstein. 2011. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." Heuristics: The Foundations of Adaptive Behavior. https://doi.org/10.1093/acprof:oso/9780199744282.003.0002.

Gilchrist, Alan L. 1977. "Perceived Lightness Depends on Perceived Spatial Arrangement." Science 195 (4274): 185–87. https://doi.org/10.1126/science.831266.

Goldreich, Oded. 2008. Computational Complexity: A Conceptual Perspective. Cambridge University Press.

Goodman, Noah A., Joshua B. Tenenbaum, Jacob Feldman, and Thomas L. Griffiths. 2008. "A Rational Analysis of Rule-Based Concept Learning." Cognitive Science 32 (1): 108–54. https://doi.org/10.1080/03640210701802071.

Goodman, Noah D, Joshua B Tenenbaum, and Tobias Gerstenberg. 2014. "Concepts in a Probabilistic Language of Thought."

Gopnik, Alison, David M. Sobel, David Danks, Clark Glymour, Laura E. Schulz, and Tamar Kushnir. 2004. "A Theory of Causal Learning in Children: Causal Maps and Bayes Nets." Psychological Review 111 (1): 3–32. https://doi.org/10.1037/0033-295X.111.1.3.

Green, E. J., and Gabriel Oak Rabin. 2019. "Use Your Illusion: Spatial Functionalism, Vision Science, and the Case against Global Skepticism." Analytic Philosophy 61 (4): 345–78. https://doi.org/10.1111/phib.12163.

Green, E.J. 2017. "On the Perception of Structure." Nous.

Green, E.J. 2020. "The Perception-Cognition Border: A Case for Architectural Division." The Philosophical Review.

Griffiths, Thomas L, Falk Lieder, and Noah D. Goodman. 2015. "Rational Use of Cognitive Resources: Levels of Analysis between the Computational and the Algorithmic." Topics in Cognitive Science 7 (2): 217–29. https://doi.org/10.1111/tops.12142.

Hafri, Alon, and Chaz Firestone. 2021. "The Perception of Relations." Trends in Cognitive Sciences 25 (6): 475–92. https://doi.org/10.1016/j.tics.2021.01.006.

Hafri, Alon, Anna Papafragou, and John C. Trueswell. 2013. "Getting the Gist of Events: Recognition of Two-Participant Actions from Brief Displays." Journal of Experimental Psychology: General 142 (3): 880–905. https://doi.org/10.1037/a0030045.

Hansen, Thorsten, Maria Olkkonen, Sebastian Walter, and Karl R. Gegenfurtner. 2006. "Memory Modulates Color Appearance." Nature Neuroscience 9 (11): 1367–68. https://doi.org/10.1038/nn1794.

Harding, Glen, Julie M. Harris, and Marina Bloj. 2012. "Learning to Use Illumination Gradients as an Unambiguous Cue to Three Dimensional Shape." PLoS ONE 7 (4). https://doi.org/10.1371/journal.pone.0035950.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. The Elements of Statistical Learning: Data Mining, Infernece, and Prediction. 2nd ed. Springer.

Hinton, Geoffrey E., and T. J. Sejnowski. 1986. "Learning and Relearning in Boltzmann Machines." In Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, edited by D E Rumelhart and J L McClelland.

Icard, Thomas F. 2018. "Bayes, Bounds, and Rational Analysis." Philosophy of Science 85 (1): 79–101. https://doi.org/10.1086/694837.

Jacob, Georgin, R. T. Pramod, Harish Katti, and S. P. Arun. 2021. "Qualitative Similarities and Differences in Visual Object Representations between Brains and Deep Networks." Nature Communications 12 (1): 1–14. https://doi.org/10.1038/s41467-021-22078-3.

Jenkin, Zoe. 2020. "The Epistemic Role of Core Cognition." The Philosophical Review 129 (2): 251–98. https://doi.org/10.1215/00318108-8012850.

Karpathy, Andrej. 2021. "Workshop on Autonomous Driving." In Computer Vision and Pattern Recognition CVPR. https://www.youtube.com/watch?v=g6bOwQdCJrc.

Kemp, Charles, Patrick Shafto, and Joshua B. Tenenbaum. 2012. "An Integrated Account of Generalization across Objects and Features." Cognitive Psychology 64 (1–2): 35–73. https://doi.org/10.1016/j.cogpsych.2011.10.001.

Kim, Junkyung, Matthew Ricci, and Thomas Serre. 2018. "Not-So-CLEVR: Learning Same-Different Relations Strains Feedforward Neural Networks." Interface Focus 8 (4). https://doi.org/10.1098/rsfs.2018.0011.

Kirchner, Holle, and Simon J. Thorpe. 2006. "Ultra-Rapid Object Detection with Saccadic Eye Movements: Visual Processing Speed Revisited." Vision Research 46 (11): 1762–76. https://doi.org/10.1016/j.visres.2005.10.002.

Kok, Peter, Janneke F M Jehee, and Floris P. de Lange. 2012. "Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex." Neuron 75 (2): 265–70. https://doi.org/10.1016/j.neuron.2012.04.034.

Konrad, Koerding, and Wolpert Daniel. 2004. "Bayesian Integration in Sensorimotor Learning." Nature 427 (January): 1–4. papers2://publication/uuid/14D22FBF-5F45-4824-9D08-9D8C52D3B47A.

Körding, Konrad P., Ulrik Beierholm, Wei Ji Ma, Steven Quartz, Joshua B. Tenenbaum, and Ladan Shams. 2007. "Causal Inference in Multisensory Perception." PLoS ONE 2 (9). https://doi.org/10.1371/journal.pone.0000943.

Körding, Konrad P., Shih Pi Ku, and Daniel M. Wolpert. 2004. "Bayesian Integration in Force Estimation." Journal of Neurophysiology 92 (5): 3161–65. https://doi.org/10.1152/jn.00275.2004.

Körding, Konrad P., and Daniel M. Wolpert. 2006. "Bayesian Decision Theory in Sensorimotor Control." Trends in Cognitive Sciences 10 (7): 319–26. https://doi.org/10.1016/j.tics.2006.05.003.

Kwisthout, Johan. 2011. "Most Probable Explanations in Bayesian Networks: Complexity and Tractability." International Journal of Approximate Reasoning 52 (9): 1452–69. https://doi.org/10.1016/j.ijar.2011.08.003.

Lake, Brenden M., Tomer D. Ullman, Joshua B. Tenenbaum, Samuel J. Gershman, Tomer D. Ullman, Joshua B. Tenenbaum, Samuel J. Gershman, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. 2016. "Building Machines That Learn and Think like People." Behavioral and Brain Sciences 40 (2017): 1–101. https://doi.org/10.1017/S0140525X16001837.

Levy, Roger, Florencia Reali, and Thomas L Griffiths. 2009. "Modeling the Effects of Memory on Human Online Sentence Processing with Particle Filters." In Proceedings of the 22nd Conference on Neural Information Processing Systems (NIPS).

Lewis, Richard L, Andrew Howes, and Satinder Singh. 2014. "Computational Rationality: Linking Mechanism and Behavior through Bounded Utility Maximization." Topics in Cognitive Science 6 (2): 279–311. https://doi.org/10.1111/tops.12086.

Lieder, Falk, and Thomas L. Griffiths. 2019. "Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources." Behavioral and Brain Sciences, 2019. https://doi.org/10.1017/S0140525X1900061X.

Little, Patrick C, and Chaz Firestone. 2021. "Physically Implied Surfaces." Psychological Science.

Lupyan, Gary. 2017. "Changing What You See by Changing What You Know: The Role of Attention." Frontiers in Psychology 8 (MAY): 1–15. https://doi.org/10.3389/fpsyg.2017.00553.

Ma, Wei Ji. 2010. "Signal Detection Theory, Uncertainty, and Poisson-like Population Codes." Vision Research 50 (22): 2308–19. https://doi.org/10.1016/j.visres.2010.08.035.

MacPherson, Fiona. 2012. "Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism." Philosophy and Phenomenological Research 84 (1): 24–62. https://doi.org/10.1111/j.1933-1592.2010.00481.x.

Mandelbaum, Eric. 2017. "Seeing and Conceptualizing: Modularity and the Shallow Contents of Perception." Philosophy and Phenomenological Research, 1–17. https://doi.org/10.1111/phpr.12368.

Marcus, Gary. 2020. "The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence." 2020. http://arxiv.org/abs/2002.06177.

Marr, D. 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. MIT Press. https://doi.org/10.1016/0022-2496(83)90030-5.

Marti, Sébastien, and Stanislas Dehaene. 2017. "Discrete and Continuous Mechanisms of Temporal Selection in Rapid Visual Streams." Nature Communications 8 (1). https://doi.org/10.1038/s41467-017-02079-x.

Michotte, Albert, G O Thine, and Geneviève Crabbé. 1964. Les Complements Amodaux Des Structures Perceptives. Publications universitaires.

Morgenstern, Yaniv, Richard F. Murray, and Laurence R. Harris. 2011. "The Human Visual System's Assumption That Light Comes from above Is Weak." Proceedings of the National Academy of Sciences of the United States of America 108 (30): 12551–53. https://doi.org/10.1073/pnas.1100794108.

Nanay, Bence. 2018. "The Importance of Amodal Completion in Everyday Perception." I-Perception 9 (4). https://doi.org/10.1177/2041669518788887.

Olkkonen, Maria, Thorsten Hansen, and Karl R Gegenfurtner. 2008. "Color Appearance of Familiar Objects: Effects of Object Shape, Texture, and Illumination Changes." Journal of Vision 8 (5): 1–16. https://doi.org/10.1167/8.5.13.

Park, Jaewoo, and Murali Haran. 2018. "Bayesian Inference in the Presence of Intractable Normalizing Functions." Journal of the American Statistical Association 113 (523): 1372–90. https://doi.org/10.1080/01621459.2018.1448824.

Phillips, Ben. 2019. "The Shifting Border Between Perception and Cognition." Nous 53 (2): 316–46. https://doi.org/10.1111/nous.12218.

Pinker, S. 1999. How The Mind Works. New York: Norton. https://doi.org/10.1111/j.1749-6632.1999.tb08538.x.

Pylyshyn, Zenon. 1999. "Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception." Behavioural and Brain Sciences 22: 341–423.

Quilty-Dunn, Jake. 2019. "Attention and Encapsulation." Mind and Language. https://doi.org/10.1111/mila.12242.

Quilty-Dunn, Jake. 2019. "Unconscious Perception and Phenomenal Coherence." Analysis 79 (3): 461–69. https://doi.org/10.1093/analys/any022.

Rahnev, Dobromir, and Rachel N Denison. 2018. "Suboptimality in Perceptual Decision Making." Behavioral and Brain Sciences 41: 1–43. https://doi.org/10.1101/060194.

Rock, Irvin. 1983. The Logic of Perception. MIT Press.

Scholl, Brian J., and Patrice D. Tremoulet. 2000. "Perceptual Causality and Animacy." Trends in Cognitive Sciences.

Schulz, Laura. 2012. "Finding New Facts; Thinking New Thoughts." In Rational Constructivism in Cognitive Development, edited by Tamar Kushnir and Fei Xu, 269–94. Elsevier.

Schulz, Laura E., and Alison Gopnik. 2004. “Causal Learning Across Domains.” Developmental Psychology 40 (2): 162–76. https://doi.org/10.1037/0012-1649.40.2.162.
Siegel, Susanna. 2017. “How Is Wishful Seeing like Wishful Thinking?” Philosophy and Phenomenological Research.
Siegel, Susanna. 2012. “Cognitive Penetrability and Perceptual Justification.” Nous 46 (2): 201–22. https://doi.org/10.1111/j.1468-0068.2010.00786.x.
Siegel, Susanna. 2017. “How Is Wishful Seeing like Wishful Thinking?” Philosophy and Phenomenological Research.
Silins, Nicholas. 2016. “Cognitive Penetration and the Epistemology of Perception.” Philosophy Compass 11 (1): 24–42. https://doi.org/10.1111/phc3.12292.
Simon, Herbert Alexander. 1997. Models of Bounded Rationality: Empirically Grounded Economic Reason. Vol. 3. MIT press.
Sipser, Michael. 2013. Introduction to the Theory of Computation. Cengage Learning.
Sokal, A. 1997. “Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms.” In Functional Integration, 131–92. Springer. https://doi.org/10.1007/978-1-4899-0319-8_6.
Spelke, Elizabeth S., and Katherine D. Kinzler. 2007. “Core Knowledge.” Developmental Science 10 (1): 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x.
Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2020. “Energy and Policy Considerations for Deep Learning in NLP.” In ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 3645–50. https://doi.org/10.18653/v1/p19-1355.
Szymanik, Jakub, and Rineke Verbrugge. 2018. “Tractability and the Computational Mind.” In The Routledge Handbook of the Computational Mind, 339–54. Routledge.
Tenenbaum, JCoshua B, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. 2011. “How to Grow a Mind: Statistics, Structure, and Abstraction.” Science 331 (6022): 1279–85. https://doi.org/10.1126/science.1192788.
Thorpe, Simon, Denis Fize, and Catherine Marlot. 1996. “Speed of Processing in the Human Visual System.” Nature 381 (6582): 520–22. https://doi.org/10.1038/381520a0.
Tokunaga, Rumi, and Alexander D. Logvinenko. 2010. “Material and Lighting Hues of Object Colour.” Ophthalmic and Physiological Optics 30 (5): 611–17. https://doi.org/10.1111/j.1475-1313.2010.00733.x.
Tooby, John, and Leda Cosmides. 1992. “The Psychological Foundations of Culture.” In The Adapted Mind: Evolutionary Psychology and the Generation of Culture., 19–136. New York, NY, US: Oxford University Press.
Turing, Alan. 1936. “On Computable Numbers, With an Application to the Entscheidungsproblem.” Proceedings of the London Mathematical Society. http://draperg.cis.byuh.edu/archive/winter2014/cs320/Turing_Paper_1936.pdf.
Ullman, Tomer D., Noah D. Goodman, and Joshua B. Tenenbaum. 2012. “Theory Learning as Stochastic Search in the Language of Thought.” Cognitive Development 27 (4): 455–80. https://doi.org/10.1016/j.cogdev.2012.07.005.
Ullman, Tomer D., Elizabeth Spelke, Peter Battaglia, and Joshua B. Tenenbaum. 2017. “Mind Games: Game Engines as an Architecture for Intuitive Physics.” Trends in Cognitive Sciences 21 (9): 649–65. https://doi.org/10.1016/j.tics.2017.05.012.
Ullman, Tomer D., Andreas Stuhlmüller, Noah D. Goodman, and Joshua B. Tenenbaum. 2018. “Learning Physical Parameters from Dynamic Scenes.” Cognitive Psychology 104: 57–82. https://doi.org/10.1016/j.cogpsych.2017.05.006.
Beers, Robert J. Van, Anne C. Sittig, and Jan J. Denier Van Der Gon. 1999. “Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model.” Journal of Neurophysiology 81 (3): 1355–64. https://doi.org/10.1152/jn.1999.81.3.1355.
Rooij, Iris van. 2008. “The Tractable Cognition Thesis.” Cognitive Science 32 (6): 939–84. https://doi.org/10.1080/03640210801897856.

Weiss, Yair, Eero P. Simoncelli, and Edward H. Adelson. 2002. “Motion Illusions as Optimal Percepts.” Nature Neuroscience 5 (6): 598–604. https://doi.org/10.1038/nn858.
Wilder, John D., Wendy J. Adams, and Richard F. Murray. 2019. “Shape from Shading under Inconsistent Illumination.” Journal of Vision 19 (6): 1–15. https://doi.org/10.1167/19.6.2.
Wolpert, David H, and William G Macready. 1997. “No Free Lunch Theorems for Optimization.” IEEE Transactions on Evolutionary Computation 1 (1): 67–82. https://doi.org/10.1109/4235.585893.
Wu, Jiajun, Ilker Yildirim, J.J. Lim, W.T. Freeman, and J.B. Tenenbaum. 2015. “Galileo : Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning.” Advances in Neural Information Processing Systems 28 (NIPS 2015), 1–9.