SIGNIFICANCE OF NEURAL NOISE

Gašper Štukelj

Graduate School of
Systemic Neurosciences
LMU Munich

Dissertation der

Graduate School of Systemic Neurosciences der

Ludwig-Maximilians-Universität München

April, 2020

Supervisors:

Prof. Dr. Stephan Hartmann

Munich Center for Mathematical Philosophy (MCMP)

Ludwig-Maximilians-Universität München

Prof. Dr.-Ing. Stefan Glasauer

Institute of Medical Technology

Brandenburg University of Technology (BTU) Cottbus - Senftenberg

First Reviewer: Prof. Dr. Stephan Hartmann

Second Reviewer: Prof. Dr.-Ing. Stefan Glasauer

External Reviewer: Dr. Andrea Loettgers

Date of Submission: 20.04.2020

Date of Defense: 01.09.2020

"What then is time? If no one asks me, I know; if I want to explain it to a questioner, I do not know"

– *Augustine of Hippo*

Summary

Two closely related analogies have proven to be central to theorizing in neuroscience and cognitive sciences, broadly construed. More precisely, reasoning about different neural mechanisms and functions they perform is very often underlined by a comparison between brains and computers, or brains and electrical circuits. This hasn't been done only to the effect of designing experiments by guiding the way neural tissues are manipulated and probed for data, but has had a profound effect on how we talk about and interpret brain activity and related behavior. The history of science is filled with fruitful, yet inaccurate analogies. Thus, an appropriate question to ask is whether some such assumptions borrowed from the engineering sciences need to be revised. This issue has been raised in particular in relation to the notion of neural noise.

A common denominator of the different definitions of neural noise has been its lack of functional importance. Neural noise, much like "noise" in computers and electrical circuits is by definition something that does not carry information. Moreover, as a rule of thumb noise is not only seen as something that doesn't contribute to the functioning of the system but even pulls in the other direction, acting as a disturbance, hindering the performance. This view is paradigmatic to engineering sciences, but has been questioned repeatedly in the last decade by both experimentalists and theoreticians in neuroscience and broader biological sciences. Indeed, the growing body of evidence suggests that much of what has historically been described and dismissed as (neural) noise in fact plays an important role in functioning of neurobiological systems.

This thesis aims at complementing these efforts of doing justice to the role of noise as a physical quantity worthy of investigation. This is done by demonstrating the central role the *notion* of noise plays in explanations and arguments commonly found in neuroscience and related fields (such as neurophilosophy and philosophy of neuroscience). With respect to the overarching topic of the thesis, the first manuscript discusses the notion of noise as a unifying concept between three behavioral domains (motor control, perception, and economic choice) which warranted the fruitful exchange of mathematical models and experimental designs between the initially separated research fields. Moreover, I argue that neural noise is an important factor that needs to be taken into account when discussing evolutionary forces that shaped decision making capacities in humans and other animals. However, much of the psychological literature has ignored its role and has rather emphasized the environment as the predominating, if not sole source of uncertainty. The second manuscript deals with a specific philosophical argument resting on the premise that brains compute with continuous signals. This premise is first rejected solely on empirical grounds. Importantly, I also argue against its conceptual fruitfulness. One of the main reasons for the latter being that the premise seems to fail to account for the way the notions of neural noise and miscomputation are used in neuroscientific explanations. The last manuscript presents a case study of neuromorphic electronics. In discussing of what makes the engineered analogs of brains computers, I argue that the methodological reason for the oversight of the importance of neural noise lies at least partially in assuming that computers can be exhaustively explained solely in terms of input-output relations. The current research on neuromorphic electronics is also interesting in that it is an active engineering field characterized by the lack of what was stereotypified as "engineering attitude towards noise".

Collectively, the three manuscripts offer a much broader take on the topics closely related to the notion of computation, ranging from the different types of computers and their differences, all the way to the computational complexity. The upshot of the thesis is that clarifying the notion of (neural) noise is of conceptual significance for many of these related debates.

Contents

List of Abbreviations

ADC	Analog to Digital Converter
AI	Artificial Intelligence
AP	Action Potential
ATP	Adenosine Triphospate
CPU	Central Processing Unit
DAC	Digital to Analog Converter
EIP	Elementary Information Process
FFH	Fast and Frugal Heuristics
FH	Fluency heuristic
GOFAI	Good Old Fashioned AI
I/O	Input/Output
LTD	Long Term Depression
LTP	Long Term Potentiation
NE	Neuromorphic Electronics
PIM	Processing In Memory
RF	Representational Format
RH	Recognition heuristic
RM	Representational medium
SAT	(constraint) SATisfaction
SIA	Selective Integration Algorithm
SSM	Sequential Sampling Model
STDP	Spike Timing Dependent Plasticity
tDCS	transcranial Direct Current Stimulation
TTB	Take The Best (heuristic)
(a)VLSI	(analogue) Very Large Scale Integration
(v)mPFC	(ventro-)medial Prefrontal Cortex
WMC	Working Memory Capacity

List of Figures

1 Introduction

In the past decade neural noise has re-emerged as a topic of interest in the neuroscientific literature. This thesis includes three manuscripts (Chapters 2–4) all of which are explicitly and rather straightforwardly linked under the broader topic of "(neural) computation". Nonetheless, the notion of noise has made its way into each one of them and plays a notable role in their respective narratives. Admittedly, one could construe my arguments concerning noise as just straightforward refinements/iterations of current debates. However, I argue that these refinements, however straightforward, lead to novel insights.

1.1 General introduction & motivation

Neuroscience is a relatively young scientific field. Nonetheless, exploring its "short" history is a very rewarding endeavor for philosophers and historians of science, and the neuroscientists themselves alike. A historic perspective is characteristic of a good deal of the research presented herein. Knowing how the use of a certain concept has changed in time and through adoptions across different research programs, is often crucial in attempting to answer more conceptual questions pertaining to scientific practice. Conversely, to make sense of such conceptual questions it is often sufficient, if not necessary, to look at the "current best scientific theories" and the respective empirical data.

Given the specific topic of scientific practice and current best theories in neuroscience and related fields, the corresponding labels of *philosophy of neuroscience*, and *neurophilosophy*, respectively, seem to describe well the different (philosophical) inquiries reported in the following chapters. Assuming a historical perspective is one of the common approaches in philosophy of (neuro)science and it is also from where I'm drawing the initial motivation.

As it is common in science, neuroscientists often describe or reason about their object of study using analogies. Loosely speaking this thesis is about multiple such analogies common to the neuroscientific literature, how they've been used throughout the development of the field, and of their present relevance. Some of these are mentioned only once, while others are discussed at length at various occasions.

There is one analogy that is clearly prevalent – namely an analogy between brains and computers. This should hardly come as a surprise. It seems that at least superficially, brains are properly described *as if* they're computing. Explicating the meaning behind the ubiquitous phrase "brains compute ..." has become the bread and butter of many philosophers and other theoreticians.

This question falls under the broader research program of the philosophy of physical computation. One of the central goals of this branch of philosophy is to delineate the group of physical objects that compute from those that don't. Therefore the first question usually comes bundled together with a second one – are brains *actually* computing? In other words, are brains computers and if so what *kind* of computers are they?

The most obvious strategy to answer both of the questions at once is to compare brains to a physical system that is *known* to compute. Indeed, brains have been compared to computing machines from the early onset of the technology (McCulloch and Pitts, 1943; Turing, 1992; von Neumann, 1951). Traditionally and independent from this debate, two different *kinds* of computers have been considered – digital and analog. Whereas some have proposed that neural systems constitute a third kind of computer (Piccinini and Bahar, 2013), others have argued that brains fall under one of the already established categories (Neumann, 1958; Beebe, 2018; Maley, 2011; Shagrir, 2010).

Computers are just one kind of electronic device, and the analogy between brains and computers can be seen as a refinement of a broader and much less controversial analogy between neurological networks and electrical circuits. The latter has only recently gained attention from philosophers interested in the history of ideas in neuroscience (Chirimuuta, 2017), leaving many topics undiscussed. The literature pool becomes larger when considering the somewhat broader topic of "the engineering paradigm", that is, the adoption of various engineering concepts and methods in other sciences. In particular, the debate on the scientific practices in the field of synthetic biology has covered a lot of ground by discussing the adoption of engineering concepts such as reliability, simplicity, robustness, and noise in the scientific discourse (Boon, 2017; Knuuttila and Loettgers, 2014, 2013).

All of the mentioned topics are addressed at various places throughout the following chapters. Nonetheless, the chapters follow each other in a sequence more closely related to the analogy between brains and computers. The first manuscript, Chapter 2, investigates to what extent does the original decision-theoretic characterization of heuristics in GOFAI still influence the conceptualization and understanding of (human) decision-making in modern research. I argue that many of the assumptions found in the present-day literature need to be revised, due to their being apparently derived from an analogy between brains and digital computers (or rather stored-program computers, see discussion in Chapter 4). Basing my arguments on both behavioral and neuroscientific data, I conclude that there's a principal difference between the latter two kinds of a (computing) physical system.

In the manuscript that follows, Chapter 3, I pick up the discussion from this general conclusion and focus on a specific proposal regarding the characterization of brains as computational systems. Specifically, I look at an argument due to Maley (2011) from which he concludes that brains are *analog* computers. While I agree with Maley's broader conclusions, I argue against both the empirical plausibility of the argument's premises as well as the fruitfulness of the proposed framework that could follow from Maley's conclusion, had the premises been accepted.

In the third manuscript, Chapter 4, I consider a different account of analog computers that I derive through a "rational reconstruction" of the early scientific and engineering endeavor behind the conception of neuromorphic electronics (NE). These devices provide an interesting and presumably less controversial token of computing physical systems that can be used for reasoning about, and comparing different (philosophical) accounts of (physical) analog computation. The closing discussion of the last chapter conveniently returns to the topics discussed in the introductory sections on (neural) noise and thus properly concludes the chapter and the dissertation itself.

While the principal questions differ between the individual chapters and some of the conclusions I've argued for are only loosely related, a significant part of the argumentative work in each chapter is carried by considerations of neural noise. It is thus the notion of noise that will serve as the main narrative thread for the rest of the discussion. The next section lays the groundwork by introducing the concept of (neural) noise and some of the related conceptual questions, followed by the discussion of how the notion of noise applies specifically to individual chapters.

1.2 Varieties of noise

In neuroscience, noise is usually considered as a *source of variability* of a signal that *is not* a consequence of deterministic properties of the system producing it (Faisal et al., 2008, p. 292). If signal is understood as a quantity carrying information, then noise might be said to be its "meaningless" component. However, it would seem that identifying something as a signal is context-dependent and would require some kind of an intentional attitude. Surely this is something that we wouldn't readily ascribe to a neuron (*e.g.*, when considering neuron to neuron communication). A similar worry might arise around a discussion of "noise management" in nervous systems, particularly, strategies involving "prior knowledge":

> By using prior knowledge about the expected structure, sensory processing can compensate for noise. This is manifest in the notion that a neuron's receptive field tells us what message the neuron is conveying. [...] Thus, the structures of receptive fields embody prior knowledge about the expected inputs and thereby allow neurons to attenuate the impact of noise. (Faisal et al., 2008, p. 298)

Arguably this issue is omitted if we simply speak of noise as "random fluctuations" (McDonnell and Ward, 2011, p. 415) or explicitly reject a factual distinction between signals and noise by conceptualizing the latter as *a type* of signal, "the value of which at any given time is drawn randomly from some distribution" (Ermentrout et al., 2008, p. 428). I'll further discuss this topic later in this section and in Chapter 4. As already demonstrated by this small literature sample, the term "noise" carries many meanings and connotations, and is thus met with various scientific or epistemological attitudes.

Taking cue from the aforementioned literature on synthetic biology, and especially from *Varieties of noise: Analogical reasoning in synthetic biology* by Knuuttila and Loettgers (2014), there are several general observations regarding the conception of noise that can be expected to hold true for neuroscience as well. In an attempt to systematize the observations and for the purposes of the future discussion I'll consider four dichotomies applying to the adoption of the concept of noise in neuroscience. Framing this a bit more generally, the dichotomies could be extrapolated to different categories, combinations of which constitute different types of analogical reasoning involving the notion of noise.[1] The distinctions will be made between negative and positive, applicative and descriptive, mechanistic and functional, and factual and conceptual analogies.

Stereotypically, analogical reasoning starts with an observation of one or more similar properties of two given objects, followed by an assertion of another similarity that has not yet been observed. Conceptualization of noise as a disruptive and detrimental component, and thus as a quantity that needs to be minimized, is one such *positive analogy* between neural circuits and engineered electrical circuits:

> On the one hand, one can pursue the positive analogy between artificial and biological systems by treating the fluctuations as a disturbance and trying to find ways of making the system more robust by changing its architecture. This approach is chosen by the engineering-oriented branch of synthetic biology, which uses different strategies to isolate and eliminate the various sources of noise.

[1] I am purposely omitting a discussion of the existing philosophical literature on analogical reasoning *as such* for a very simple reason (although see Chapter 4). The literature is too broad and even a short review would be well beyond the scope of this thesis. The list of dichotomies could be thus seen as a mere narrative tool for introducing the ways the concept of noise is used in neuroscience relevant to the specific examples discussed in the next chapters.

> (Knuuttila and Loettgers, 2014, p. 86)

This brings us to the first dichotomy, namely between "positive" and "negative" analogies. Whereas noise has been conventionally viewed as a disturbance or a disruptive force that is limiting the extent of control, there is a growing literature demonstrating its various benefits for a properly functioning biological system.

> The basic science approach, by contrast, has chosen the opposite direction, drawing a further negative analogy to artificial control systems. Recognizing noise as an intrinsic part of biological systems, the researchers in this field have started to study the sources and impact of noise on biological systems. As a result of these studies, noise has also been assigned a functional role; it supports the various functions of biological systems. (Knuuttila and Loettgers, 2014, p. 86)

Notably, Knuuttila and Loettgers (2014) consider the division between the use of positive and negative analogies between biological and artificial systems to be more or less aligned with the separation between adopting either an "engineering-oriented"[2] or a "basic science" approach to the research topic.

> This [application-oriented] branch of synthetic biology does not aim to mimic biological systems but to engineer novel systems with specific functions, which need not be brought about in the same ways

[2] The term "engineering approach" might be a bit of a misnomer, as noise is not always conceptualized as a disturbance by engineers. For example, "process noise" in the Kalmann filter simply describes the innate variability of the state, independent of external perturbations.

> as in naturally evolved systems. Because of this goal, and also due to the close ties with engineering, noise is predominantly regarded as a disturbance within this branch, to the extent that it reduces control over the designed biological systems. Much effort has therefore been invested in strategies to avoid or reduce noise. (Knuuttila and Loettgers, 2014, p. 81–2)

As discussed later in Chapter 4, some engineering projects related to neuroscience actually maintain a relationship of a positive analogy with their "target" biological systems by means of "forcing" a negative analogy with respect to other engineering paradigms. In addition to a dichotomy between negative and positive analogies, one might thus find it useful to also talk separately about "applicative" and "descriptive" uses of an analogy.

The discrepancy between engineered electrical artifacts and biologically evolved neural networks has also been discussed in neuroscientific literature in relation to conceptualizing noise in terms of an information carrying or, at the very least, signal enhancing quantity. The term "stochastic resonance" was initially coined to describe a very specific phenomena, but with time it became a kind of a catchall phrase for a myriad of observed benefits of noise. Importantly, McDonnell and Ward (2011) argued that a more detailed conceptual framework is needed for a proper grounding of the emerging field of neural noise research. More specifically, their proposed term "stochastic facilitation" is meant to generalize the notion of stochastic resonance while also breaking away from the historical and conceptual baggage carried by its predecessor. One of the conceptual reasons they offer in favor of adopting the new terminology is particularly illustrative of another important distinction regarding analogies involving a notion of (neural) noise:

> Notwithstanding the above semantic issues, stochastic resonance

> stands apart from other identified constructive roles for noise in that all existing definitions require identification of an input signal and an output signal. This immediately associates the concept with notions of information processing and computation, as in engineered signal processing systems. Consequently, stochastic resonance is often described as paradoxical or counter-intuitive, because in engineered electronic systems noise is naturally seen to be only detrimental to quality. However, in a biological context, the effect is hardly counter-intuitive when thought of as the benefits of randomness, as with other constructive roles of noise in which inputs and outputs need not be readily identifiable. (McDonnell and Ward, 2011, p. 417)

The distinction between framing the benefits of noise in either information-theoretic or physiological terms is reminiscent of the broader philosophical debate on the nature of explanations in computational neuroscience. While some authors have stressed the importance of "efficient coding" hypotheses or other computational frameworks (Chirimuuta, 2014; Eliasmith, 2010), others have claimed that proper explanations necessarily involve an identification of a mechanism bringing about the explained phenomena (Kaplan, 2011; Kaplan and Craver, 2011). While the debate is beyond the scope of this introduction, I will borrow some of the terminology when drawing a distinction between functional and mechanistic analogies involving neural noise.

> It is likely that observations of stochastic facilitation in the brain can be explained in terms of the randomness arising from stochastic biological noise enabling the operation of a mechanism that implements a computational task. Clearly, there could be a diverse range of neural mechanisms in which this dependence of 'algorithm imple-

> mentation' on noise could occur. (McDonnell and Ward, 2011, p. 419)

It is important to be precise about what this dichotomy actually delineates. For example, Faisal et al. (2008) discuss how the inevitable fact of noise has strongly shaped the way information is processed in neural systems. They list various examples of noise management strategies for mediating and sometimes also exploiting its effects (structures embodying "prior knowledge", weighted averaging, population coding, stochastic resonance, randomized state exploration, *etc.*). Notably, Faisal and colleagues also demonstrate how some *physiological* observations could be well explained by appealing to the presence of neural noise:

> In axons of less than $0.3\mu m$ diameter, the input resistance is large enough that spontaneous opening of single Na^+ channels at the resting potential can produce 'Na^+ sparks' that can trigger APs in the absence of any other inputs. These 'rogue' APs become exponentially more frequent as axon diameter decreases, rendering axons below $0.08 - 0.10\mu m$ diameter useless for communication. This lower limit matches the smallest diameters of axons across species. Analogously, noise sets the lower limit for the diameter of excitable cell bodies to $\sim 3\mu m$.

Notwithstanding the fact that this is a claim about an "implementation" detail, the analogy leading such reasoning would not fall into the mechanistic category. Note that the purported explanation of the observed physiological fact relies on a positive, descriptive, and *functional* analogy between electrical information-processing systems and brains. We expect the diameters of axons or cell bodies to be within a certain range, insofar as we expect them to perform a certain function. This kind of reasoning contrasts well with another example discussed by Knuuttila and Loettgers (2014), namely some of the arguments found in the seminal research on circadian clocks. Following a mechanistic analogy with engineered systems, biologists and neuroscientists have reasoned that the mechanisms behind the observed oscillatory behavior must be due to a underlying mechanisms including a negative feedback loop (Bechtel and Abrahamsen, 2010; Bechtel, 2012). More generally, recognizing that neural circuits need to operate stably under continuous changes and do so in absence of a central regulatory "unit" has lead to the conjecturing and later establishing the existence of different homeostatic mechanisms.

> There can be no central, global control mechanism monitoring and adjusting the properties of each individual cell in a coordinated manner. Instead, global control is observed as an emergent feature of the nervous system, arising from the combined effects of a hierarchy of regulatory mechanisms operating on the level of cellular networks, individual cells, subcellular domains and, ultimately, individual genes and proteins. (O'Leary and Wyllie, 2011, p. 4812)

There's one important caveat to the discussion examples I used above to motivate the dichotomy between analogies used to describe a function and analogies used to describe a mechanism. Prefacing it with an important observation made by McDonnell and Ward (2011) regarding the centrality of the "input-output modeling methodology" (Shagrir, 2018) is rather misleading, insofar as it doesn't only underlie the functional, but also the mechanistic approach to system description and explanation.[3] While some proponents of the

[3] Nonetheless, the distinctions are not completely disassociated. While functional or computational descriptions are trivially and necessarily entrenched in thinking in terms of inputs and outputs, mechanistic descriptions/explanations need not be.

mechanistic explanations have argued for a distinction between computing and information-processing mechanisms (Piccinini and Scarantino, 2010), the identification of I/O relationships is still an integral part of mechanistic explanations (Bechtel, 2012). In fact, "endogenous activity" of neural systems that is hard to describe in terms of I/O processing is often indiscriminately categorized as noise. As discussed by McDonnell and Ward (2011), this activity is then "surprisingly" or "paradoxically" observed to be of significant importance for the functioning of the system.

> While variability in the signal recorded is noted, it is generally treated as noise that renders it difficult to extract what is regarded as the signal that reflects the response to the stimulus. In fact, such noise often reflects the endogenous activity of the system. Far from being in a constant state, the mechanism varies over time and this has consequences for the activity that might be evoked by what are usually taken as the inputs to the mechanism. (Bechtel, 2012, p. 236)

There is one more dichotomy to consider. Most of the time when an analogy with electrical systems enters a debate in biology or neuroscience, it is used for *factual* considerations. In this case, the argument is that some *things* do (or do not)[4] perform a similar function or employ a similar mechanism as some other things. Where "things" denotes mind-independent, factual objects. On the other hand, much of the different analogies I make use of in the subsequent chapters of this thesis, pertain to *mind-dependent* objects or loosely speaking, "concepts". While the difference between positive and negative *conceptual* analogies is trivial, one would be hard-pressed to try and describe how

[4] This phrasing is related to combinations with descriptive category, phrasing of an applicative analogy could be expressed with a change in modality from "do" to "should".

the difference between functional and mechanistic analogies applies. Finally, the applicative-descriptive dichotomy seems to line-up with the distinction between observing or asserting how a certain concept is or should be used.

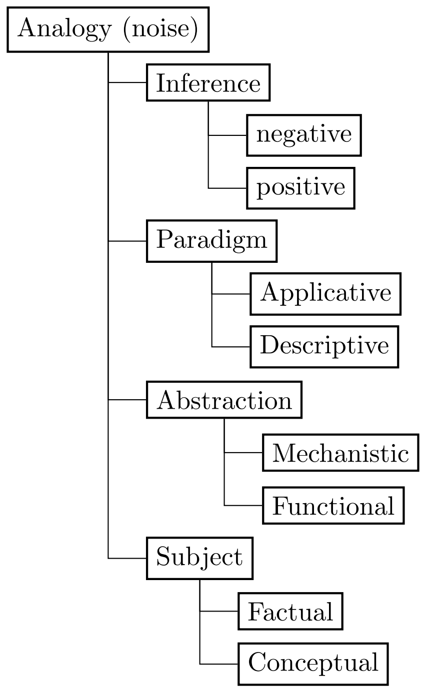

Figure 1: Classification of analogical reasoning in (biological) sciences.

A graphical summary of the four dichotomies discussed in this section can be found in Figure 1. Equipped with these labels, I'll continue by describing how I used both the notion and the fact of neuronal noise to argue about different conceptual questions regarding neuroscience and related fields in the subsequent chapters. It is these and other arguments, and reasoning patterns discussed herein that make the construct of neural noise conceptually significant.

1.3 Conceptual significance of neural noise

Making use of the discussion in (Knuuttila and Loettgers, 2014) once more, it is worth pointing out that the authors considered two separate cases of a negative analogy with the engineering sciences. More specifically, the bifurcation of synthetic biology into two distinct approaches is said to have happened with the one dubbed as the "basic-science approach" breaking with the engineering design patterns based on noise attenuation and modularity.

The assumption of modularity is discussed at length in Chapter 2 and Chapter 4. In the former I argue that the analysis of heuristics, cognitive decision-making strategies initially conceived as non-exhaustive tree search algorithms, is based on the faulty assumptions of the underlying mechanisms being both modular and sequential. As discussed by Bechtel (2012), such thinking is characteristic of mechanistic explanations and has a lot to do with I/O modeling methodology (see Section 1.2).

> Central to mechanistic explanation as it has been pursued in biology is the assumption that the behavior of mechanisms is to be understood in terms of the operations performed by their parts and that therefore it is essential to decompose mechanisms into their parts and operations. (Bechtel, 2012, p. 234)

> [R]esearch usually begins by positing the simplest arrangement in which multiple parts are organized to generate the phenomenon – a sequential arrangement in which the product of one operation is provided as an input to the next operation, which transforms it and passes it to yet another operation, as in an assembly line. (Bechtel, 2012, p. 235)

This fits well the historical narrative I adopt in Chapter 2. The positive mechanistic analogy between neural circuits underlying decision-making and digital computers, or more precisely von Neumann architecture, is what has been driving the functional analogy between computer algorithms and human behavior. Given that computers are irreducibly I/O systems it is reasonable to expect that it is due to this analogy that in the heuristics literature the endogenous activity[5] of neural circuits in question hasn't been only pushed aside as "meaningless noise", but rather ignored altogether.

The field of decision neuroscience consists of three research programs, traditionally thought to be distinct and independent: cognitive, perceptual, and motor decision-making. These terms describe behavior in value (or preference) based choice, perceptual classification, and motor-control tasks, respectively. Similarly, the heuristics literature includes many accounts of purported heuristics used for perception and some research has been done on motor control heuristics as well.

A conceptual commonality between heuristics research and decision neuroscience, as well as among the different fields within the latter, is that all three behavioral domains (economic choice, perception, and motor control) have been described in terms of probabilistic decision-making. A crucial difference lies in the heuristic literature identifying the environment as the sole source of the uncertainty, whereas the decision neuroscience literature has strongly emphasized the role of neural noise.

Thus explanations invoking the notion of neural noise are abundant in decision neuroscience literature. Specifically, by appealing to the fact of neural noise and a positive analogy with noise attenuating electrical circuits, neurosci-

[5] That is, the activity that cannot be described in terms of I/O, see (Bechtel, 2012) and the discussion above.

entists first conceived functional similarity between economic choice, perception, and motor control, which has later led to a discovery of a canonical computation, namely normalization.[6] The overemphasis of the environmental factors and a complete lack of consideration of the neural noise, especially in the later stages of a decision-making process (after all the information has already been gathered) has had a significant impact on the discussion and characterization of different decision-making strategies.

The discussion of a negative analogy between brains and digital computers continues in both Chapter 3 and Chapter 4. In the former I also consider another example of the concept of neural noise playing a central role in a neuroscientific explanation. The chapter is organized around analyzing a specific argument put forward by Maley (2011) aimed at showing that brains are not digital, but rather analog computers. More specifically, I consider various shortcomings of a premise stating that brains compute with continuous variables.

The link between neural noise and discrete computation has already been discussed by Eliasmith (2000), who argued that due to neural noise brains compute with discrete variables and should be thus considered as digital computers. I build upon a contraposition of Eliasmith's conditional – if brains are not discrete computers, then there is no neural noise. This then serves as a basis of a *modus tollens* argument against Maley (2011), insofar as he fails to provide an alternative conceptual tool with which we'd be able to make sense of common neuroscientific explanations involving notions of computational identity and miscomputation.

The notion of an analog computer is also discussed in Chapter 4 in

[6] Normalization has been first proposed as a computation performed by the visual system. It was later "discovered" that it is a canonical computation, in the sense that it can be found in different brain areas underlying distinct behavioral domains.

which I consider an applicative negative analogy between brains and von Neumann architecture that resulted in the conception of neuromorphic electronics research. I try to explicate the idea that neuromorphic computers are analog computers, which leads me to revisit some topics from Chapter 2, namely non-modularity and computational primitives. The research through a design of neuromorphic electronics has had a significant impact on our understanding of brains, and continues to do so with illuminating the different ways noise is utilized in neural computation.

2 On the simplicity of simple heuristics

Recent[7] evidence suggests that the take-the-best heuristic (TTB) – flagship of "fast and frugal heuristics" research program (FFH) – might in fact not be as frugal as tallying, which is considered to be a more complex strategy. Characterizing a simple decision strategy has always seemed straightforward, and the debate around the simplicity of the TTB is mostly concerned with a proper specification of the heuristic. I argue that the predominate conceptions of "simplicity" and "frugality" need to be revised. To this end, a number of recent behavioral and neuroscientific results are discussed. The example of TTB serves as an entry point to a foundational debate on bounded agency. I argue that the FFH needs to question some of its legacy from the classical AI research. For example, the assumption that the bottleneck of decision-making process is information processing due to its serial nature. These commitments are hard to reconcile with the modern neuroscientific view of a human decision-maker. In addition, I discuss an overlooked source of uncertainty, namely neural noise, and compare a generic heuristic model to a similar neural algorithm.

[7] This chapter is an only slightly modified version of the manuscript that was accepted for publication in *Adaptive Behavior*, see (Štukelj, 2019).

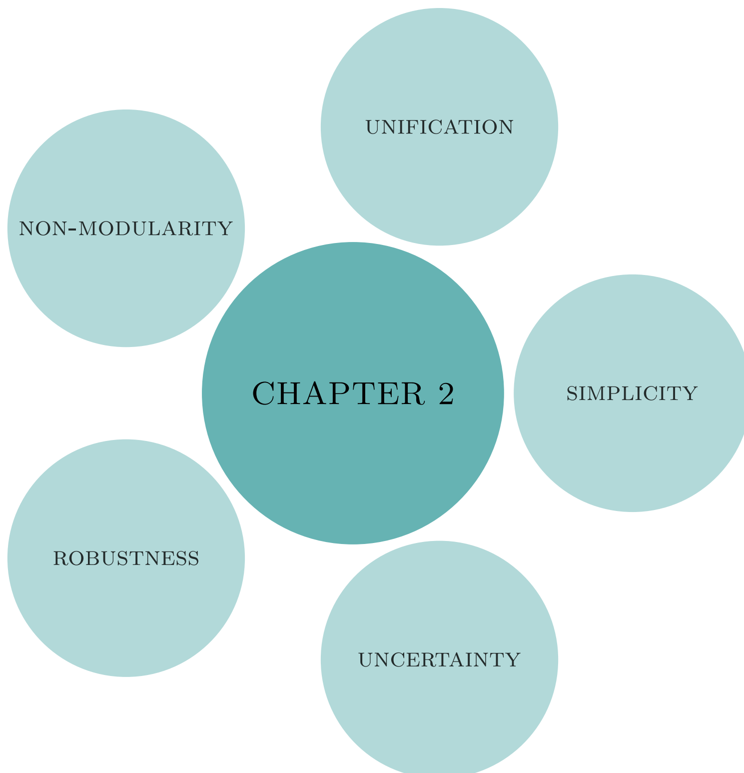

Figure 2: Graphical summary of Chapter 2.

2.1 Introduction

Bounded agency denotes judgment and reasoning under constrained resources. In order to quantify agent's problem-solving capacity we need to measure costs imposed on her as she performs a specific strategy. Depending on the chosen measure, one might give a different answer on how an agent will go about balancing the costs with the related prospect of her reaching a set goal. One particularly popular way is to define a set of "elementary information processes" (EIPs) (Simon and Newell, 1971) and then associate each EIP with a specific cost. Assuming that a strategy can be decomposed into a sequence of EIPs, its complexity can be then measured as a weighted sum of the costs of its constituents (Bettman et al., 1990). One of the most simple strategies according to this approach is the lexiographic rule.

Lexicographic rule or "take the best" (TTB) is one of the "fast and frugal" heuristics, a set of simple decision strategies proposed to explain human decision-making and judgment (Gigerenzer and Gaissmaier, 2011). TTB is characterized by ignoring all of the information beyond what is necessary to distinguish between alternative choices (see fig. 1). It searches sequentially through an ordered sequence of choice attributes until a single tuple of attribute values is found, such that one choice is preferred over the others. This apparent frugality has convinced many economists, psychologists, and philosophers to consider TTB to be one of the least effortful strategies.

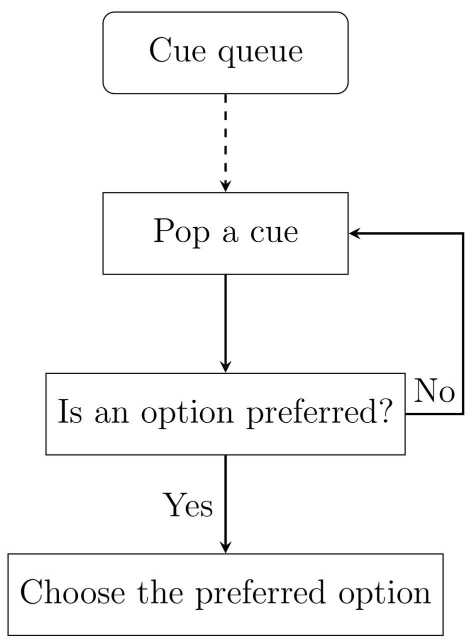

Figure 3: Take-the-best heuristic.

Simulating the use of heuristics with an artificial cognitive system, bounded by neurophysiological constraints similar to those of humans, suggested that TTB is more costly than it was previously thought (Fechner et al., 2018). The surprise came with an alternative and presumably more complex heuristic, tallying (see fig. 2), sometimes leading to faster and less effortful decisions. Whereas TTB often ignores some if not most of the cues, tallying counts them all, ignoring only the weights.[8] Since TTB also ignores weights insofar it only looks for a cue that discriminates between choice options, the number of cues examined by TTB will be at most equal to the number of cues examined by tallying. Considering the additional counter functionality, it seems that tallying consists of strictly more EIPs than TTB.

[8] Differences in cue weights could be important for strategy selection and an ecologically rational agent chooses the strategy fitting to the environment and her goals. However, this topic is beyond the scope and interests of this paper.

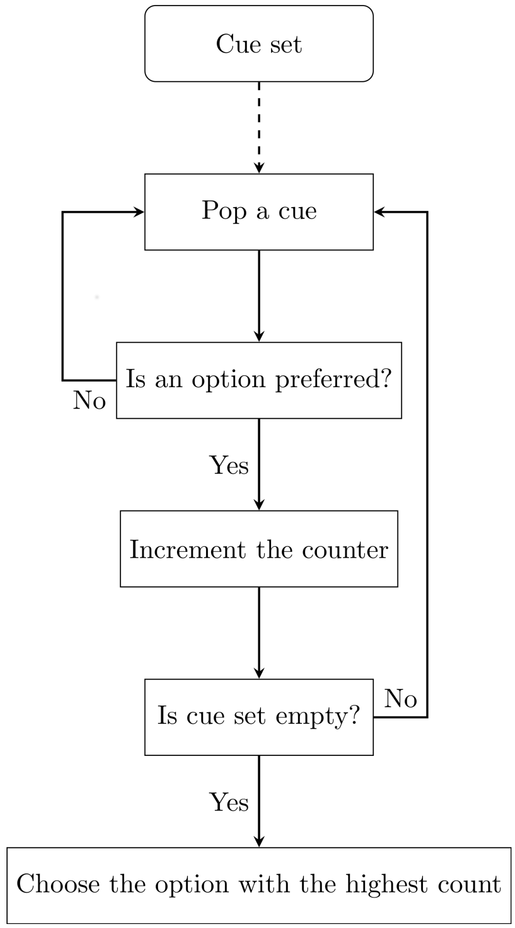

Figure 4: The tallying heuristic.

Observing under which experimental conditions the results were replicated in human behavior, researchers argued that the *in silico* and *in vivo* observations all go to show that this is due to the information having to be ordered before TTB can be applied.[9] Authors suggest that when the information is pre-

[9] I've included this assumption in the computational scheme of the TTB with a `cue queue` node indicating an ordered data structure. Similarly, using a `cue set` indicates random (unordered) accessing of information in tallying. Any other pair of an ordered and an unordered data structures would do as well.

sented randomly, ordering processes impose additional costs by further taxing an agent's working memory capacity (WMC). However, no strong conclusions can be drawn due to the fact that humans can often use appropriate robust cue orders without explicitly computing them (Katsikopoulos et al., 2010). These may come intuitively from the perceived casual structure of the environment, pattern recognition based on prior experience, by imitating other agents, or simply picking up on the natural frequencies of a cue and the target occurring together.

Moreover, there's another notion of simplicity invoked in the literature on heuristics (Gigerenzer and Brighton, 2009; Brighton, 2019; Mousavi and Gigerenzer, 2017). It was argued that heuristics draw their value from being appropriate tools to deal with "unspecifiable sets of outcomes, or unknown probabilities associated with them" (Mousavi and Gigerenzer, 2017), as opposed to situations in which probabilities and outcomes are known. In the first sense heuristics are viewed as simple algorithms, because they're built out of a small number of EIPs. In the second sense heuristics are simple because they ignore information in order to minimize preferential or inferential choice variance. Simplicity$_1$ is a target feature, as a cognizer strives towards a low consumption of computational resources (e.g. WMC). Simplicity$_2$ is an enabling feature, as it guards against overfitting in an epistemically unfriendly environment. Respectively, fast and frugal heuristics are claimed to be decision algorithms that 1) minimize computational effort and, 2) are robust tools for dealing with unquantifiable sources of uncertainty.

These two notions are logically independent.[10] And the results in (Fechner et al., 2018) don't challenge any. The authors rather argue that, under certain circumstances, tallying might be more simple than TTB insofar it would

[10] A robust algorithm can be either costly or simple, and *vice versa*.

require less computation to be performed. I argue that complementary to this conclusion at least part of the overlooked costs associated with the TTB are due to the requirement to ignore some information, challenging the view of simple heuristics minimizing computational effort. I start by reviewing direct and indirect behavioral evidence in section 2.2 suggesting that ignoring information is cognitively taxing.

The review of behavioral evidence and the latter sections overall are intertwined with a historical discussion of seminal research on bounded agency. The notion of robustness through simplicity is a foundational contribution of the ecological rationality research program. However, the idea of effort minimization through algorithmic simplicity is a tenet of its predecessor, information processing psychology program led by Herbert Simon and Allen Newell. Importantly, it had a strong influence on Simon's original conception of bounded rationality. When considering Simon and Newell's work, it becomes clear that the bounded rationality concept cannot be divorced from their research on classical AI. Thus many issues debated herein can be traced back to some legacy assumptions based on an analogy between human brains and digital computers. The central assumption being that neural architecture consists of separate processing and mnemonic modules, and is thus subject to computing bottlenecks that result in information being processed serially. It is under this assumption that certain heuristics are deemed simpler in terms of reduced computation.

Beside direct evidence of cognitive costs of ignoring information, section 2.2 revolves around two important observations. Namely, that there are strong contextual effects suggesting evidence is processed in parallel rather than sequentially, and a lack of consistent decoupling of information search and information processing, in theoretical discussions as well as in experimental designs. Both motifs appear in latter sections and prove to be crucial for a proper synthesis of various areas of research related to decision-making. The article continues with section 2.3 in which I discuss the literature on perceptuo-motor decision-making. While the detailed argument for the relevance of this research is given in the beginning of the section, it should suffice to say that both heuristics and decision neuroscience literature suggest that the same principles underly the two decision-making domains.

The section on perceptuo-motor decision-making was written with two goals in mind. First, to reinforce the argument made earlier by offering new evidence about non-seriality and a clear division between search and processing of information. Second, it leads to a discussion of an overlooked source of unquantifiable uncertainty, namely neural noise. The motivation for reviewing literature on perceptuo-motor decision-making is thus in part practical, as it overlaps greatly with the decision neuroscience literature. It bridges the discussion between two separate notions of simplicity, and segues into section 2.4 where the discussion is moved from behavioral to neural models. The motivation to do so is not lacking, as one would be hard-pressed trying to argue that neural mechanisms can be completely abstracted from when trying to understand *biological* limits of human decision-making.

> Ecological rationality is a response to the problem of understanding how biologically constrained organisms function under the uncertainty of the natural world (Todd and Brighton, 2015).

The overarching argument of this paper rests on a premise that a bounded rational agent is bounded *qua* a neurobiological system. It follows that proper understanding of the relevant neural architecture is crucial to veridical characterization of notions of bounded and hence also ecological rationality. In fact, a great deal of effort was dedicated to finding neural basis of some fast and frugal decision rules in the ecological rationality literature (Gigerenzer and Gaissmaier, 2011; Khader et al., 2016). There is also historical motivation.

Validating the theory based on the research on perception and underlying neurophysiology has already been envisioned in Simon and Newell's sketch of the research strategy Simon and Newell (1971):

> 8. Search for new tasks (e.g., perceptual and language tasks) that might provide additional arenas for testing the theories and drawing out their implications.
>
> 9. Begin to search for the neurophysiological counterparts of the elementary information processes that are postulated in the theories.

Taking cue from Simon and Newell I do just that. I argue that in lieu of sequential machines, we'd do better to think about neural architecture that is characterized by parallelism, recurrency, mutual inhibition, and non-modularity. This leads the discussion of a neural mechanism that was first discovered in the perceptual domain, but has recently been implied in cognitive decision making as well. Examining an algorithm based on this particular neural mechanism we can see many motifs discussed in decision neuroscience and bounded rationality literature coming together. Curiously enough, the neural algorithm at least superficially exhibits the same computational properties as a generic heuristic model. It seems to operate sequentially and to systematically ignore parts of information. Based on the established difference between search and computation, and the particular properties of neural computations, I argue that human decision-making is likely led by the constraints of robustness, but not computation-minimization.

2.2 Discerning search from computation

As noted by (Fechner et al., 2018) complexity analysis based on EIPs presupposes that cognitive processes are serial. If steps are performed in series, each new step adds to the overall computing time. Inversely, less steps result in a faster execution. The heuristic schema presented earlier assume that cues are evaluated one-at-a-time. Evaluating a single cue amounts to going through the loop once. For a given set of cues, the TTB will require the same amount of loops as tallying only in the worst case scenario, when none or only the last cue discriminates between the options.[11] Since on average TTB will halt after a smaller number of loops it is believed to be a faster and a more frugal strategy.

Sequential evaluation of different chunks or sources of information (cues) remains an explicit feature of a prominent class of "non-compensatory" heuristics (Gigerenzer and Gaissmaier, 2011). Speaking broadly for memory-based decisions, such models assume that choice attributes are selectively and sequentially retrieved from memory. Accordingly, the strategy that requires less information will result in a smaller number of retrievals. Two such closely related strategies are the recognition (RH) and fluency heuristics (FH), which prompt an agent to choose an option that is remembered faster or with more ease, respectively.

A recent neuro-imaging study provided contrary evidence (Khader et al., 2016). When an agent is presented with an option set, all of the associated memories, or rather their respective neural representations, are automatically activated. Only after such mass retrieval can she pick out memorized values of target attributes. However, it takes additional cognitive control and attentional focus to boost the activation of memory representations related to a specific attribute.[12] Ignoring retrieved content doesn't necessarily cut cognitive costs; even more so, it is likely to increase them! On its own, this is already a very

[11] To simplify, I assume as it is common to, that if no discriminating cue is given, an option is chosen randomly.

[12] Note that, due to metabolic efficiency, this boost is probably achieved by inhibition of attributes corresponding to the information that strategy requires to be ignored.

telling result, since the RH and FH are just two special cases of TTB, all three sharing the same computational schema.

The idea of seriality is a legacy of symbolic AI, or more precisely Simon's and Newell's conviction to a strong analogy between humans and digital computers, the "two most significant classes of [physical] symbol systems" (Simon and Newell, 1971). Classical AI research was in great part motivated by the promising insights of reasoning about the structure of software capable of replicating human behavior.

This was done for solving well structured problems, like chess, that are amenable to description in a rigid, formal language. To solve a well defined problem, rules need to be remembered and held in memory as they're applied one after another, until a solution is found. This requires sustained attention and explicit, deliberative application of the solving strategy. As such, it provides a textbook example of a task that requires intense reliance on working memory, one of the most scarce cognitive resources.

> The fact of limited resources allows us, for most purposes, to view a symbol system as though it were a serial, one-process-at-a-time device (Newell and Simon, 1976).

However, reasons for emphasizing WMC limitations and assuming that information processing is the decision process bottleneck go beyond just observational design. Asserting that something is a *physical symbol system* is to assert something about its architecture (Newell and Simon, 1976). Core architectural principle of a digital computer is modularity. Both size of a system memory, and the bandwidth between it and the processing unit remain the main tie-ups of computer processing speed to this day. This is often referred to as a "von Neumman bottleneck". Both assumptions are to at least some degree present in the modern literature as well.

> **Adaptive toolbox**: the cognitive heuristics, their building blocks (e.g., rules for search, stopping, decision), and the core capacities (e.g., recognition memory) they exploit (Gigerenzer and Gaissmaier, 2011)

While lately it is more often than not argued that ignoring information is primarily done to increase accuracy, it is still retained that being "frugal", that is, forgoing some computations, makes an algorithm faster by decreasing the cognitive effort needed for its execution. "[A] heuristic is a strategy that ignores some information and minimizes computation" (Gigerenzer and Sturm, 2011). Information processing remains construed as an effortful process constrained by a WMC. That is why it is commonly assumed that humans only process a small amount of information and systematically ignore the rest (Gigerenzer and Gaissmaier, 2011; Hertwig and Pachur, 2015; Hertwig and Engel, 2016).

> **Heuristics**: strategies that ignore information to make decisions faster, more frugally, and/or more accurately than more complex methods (Gigerenzer and Gaissmaier, 2011)

Literature on fast and frugal heuristics often conflates information search with computation. For example, (Todd and Brighton, 2015) define heuristic both as an "information-processing mechanism" and as something that is "guiding search for crucial information". In fact, this follows readily from the assumption that, amongst others, the "building blocks" of a heuristic include rules for both search and decision. This amphibious nature of heuristics is strongly reminiscent of Simon and Newell's conceptualization of computation in terms of search in a tree data structure. In fact, in cognitive science heuristics ofter assume their original role of guided tree-search algorithms (Gigerenzer et al., 2012; Fu, 2016). The difference between search and calculation points at an ambiguity of what it means to be frugal by ignoring information. An agent can either forgo searching new information, or omit computing with it, even though she already acquired it.

Some of the most pertinent objections to heuristic models have been raised by comparing them with an alternative model based on parallel processing in a recurrent winner-takes-all network of nodes representing both cues and behavioral outputs (Glöckner et al., 2014). To that end a number of behavioral protocols were developed to test divergent predictions about response times, confidence reports, eye movements, and other behavioral variables (Glöckner et al., 2014; Söllner et al., 2014; Glöckner and Betsch, 2012).

A common theme to all the arguments is a distinction between information search (acquisition) and information processing. Both decisions and confidence are influenced by irrelevant, yet valid information when it is freely perceived. It seems that humans (adaptively) ignore free information only when the act of perception is within their power. That is, sometimes they do not *acquire* additional information, even when it would come without a cost, yet they always *evaluate* all of the available cues. It was observed that removing a low-validity cue in a non-compensatory environment[13] can result in a longer reaction time, despite there being even less information to be processed.

In summary, abundant behavioral evidence suggests that humans don't ignore perceived or otherwise attained information. It was argued forcefully that deliberate control, and thus WMC, has mostly to do with the attaining of the

[13] A non-compensatory environment is characterized by great discrepancies between cues' predictive powers, for example, due to high correlation between the cues. Knowing the most predictive cue thus eliminates the need for knowing the others. It is a type of environment in which TTB is particularly successful, and arguably a prescriptive strategy of choice.

information, but not computing with it. Secondly, if cues would be processed one-by-one, we would expect the evaluation to be invariant with respect to the changes of the broader body of information. To the contrary, strong contextual effects indicate that different chunks of information are not evaluated separately. Taken together, we can see why suppressing parts of information amounts to active, and hence effortful, intervention into an arguably "holistic" process.

A growing body of behavioral evidence suggests that the decision-making bottleneck occurs at the stage of information acquisition. For example, when humans are learning a causal structure of an experiment, they often don't form correct hypotheses, but if the hypotheses are given, they have no trouble computing with the respective probabilities (Bonawitz and Griffiths L., 2010). Additional working memory load only affects memory processes when new information is being encoded, but not during retrieval of already learned information (Sprenger et al., 2011). I find this all well aligned, with the motor experiments performed by (Acerbi et al., 2014) that "suggested that suboptimality in dealing with complex statistical features [...] may be due to a problem of acquiring the priors rather than computing with them". The relevance of such experiments is discussed in the next section.

2.3 Heuristics for perception and motor-control

If heuristics are build upon evolved capacities, and are thus partially shared across species (Gigerenzer and Sturm, 2011), it would seem likely that some of these "building blocks" were also used for heuristics for other domains of human agency. Indeed, simple strategies like "gaze heuristic" have been proposed for solving perceptuo-motor tasks as well (Gigerenzer and Gaissmaier, 2011; Raab, 2017). How would an EIP analysis apply to these?

From a purely conceptual point, it is instructive to look at the early discussion of what constitutes a complex motor act (Schieber, 1990). Intuitively, hand synergies, like grasping an object or forming a fist, are composed of atomic movements, like moving a single digit. Analogous to heuristic models research, listing the latter would give us a register of "elementary motor processes". The complexity of a composite movement could then be estimated from the complexity of its constituents. Yet, for most of the digits, neural circuits implicated in individuated movements encompass circuits for composed movements that involve a given digit and recruit even additional neurons.

That is, moving a digit triggers movements of other digits as well. And this goes beyond a mere mechanical coupling of tendons. In order to move a single digit, movements of several others need to be suppressed by activating additional inhibitory neurons. That is why such intuitively elementary motor processes cost more both in terms of neuronal metabolic consumption and, until a movement is sufficiently rehearsed, effortful cognitive control.

The analogy between individuating digit movements and selecting specific decision option attributes is self-inviting, but it is probably nothing more than an analogy. Nonetheless, it is a foretoken of how misleading our intuitions can be when it comes to computations in neurobiological systems. And as we'll see, looking at a perceptuo-motor system provides well understood examples of generic neuronal computations, some of which we can also expect to underlie higher cognitive functions. To be clear, I'm certainly not the first to point at a similarity between perception, cognition, and motor control in the context of heuristics.

Just like there are perceptual illusions, it was reasoned that there have to be "cognitive illusions" as well. This symmetry argument didn't inform the theory as much as the experimental design (Kahneman, 2003). That is why it is often criticized for seeming intention to provoke non-representative reasoning patterns. More importantly, a number of insights in neuroscience came from describing perception, and motor control as a decision problem. A perceiving, or moving agent is said to be deciding between categorizations of a perceptual stimulus, or between different movement trajectories, respectively. Different categorizations (movements) can be associated with differently valued outcomes, thus giving rise to a non-trivial decision problem (Hanks and Summerfield, 2017; Wu et al., 2015).

At first it seemed that perception and motor control don't have much in common with cognitive decisions. Near optimal performance in the perceptual and motor decision tasks was often contrasted to seemingly irrational behavior in classical economic paradigms (Summerfield and Tsetsos, 2015). However, a great deal of recent literature argues that there is more between the perceptual and the economic decision making than just normative and nominative commonalities. This is mostly due to the development of shared decision models and mechanistic explanations, and experimental evidence implying overlapping brain regions and similar physiological signatures (Hanks and Summerfield, 2017; Philiastides et al., 2010; Polanía et al., 2014a). With more and more interest in comparing perception and economic choice, researchers developed protocols that use the identical stimuli, or even the same stimulus property (Polanía et al., 2014a; Dutilh and Rieskamp, 2016).

The limits of human decision-making have been primarily explained with the putatively high cognitive costs of information processing. With the implication that a bounded rational agent inevitably ignores some information due to a processing overhead, and possibly bettering her decisions in the process. A notable amount of evidence was gathered against this behavioral prediction, both in economic and perceptuo-motor decision literature. In classical cued decision-making under uncertainty humans ignore information sources when exploring a problem space, but do not ignore freely given information (Glöckner and Betsch, 2012). It was suggested that cognitive limitations, such as WMC, have mostly to do with acquiring information, but not manipulating it (Glöckner et al., 2014).

Similar observations have been reported in perceptuo-motor tasks as well. Experiments suggest that all of the available information is utilized, resulting in strong contextual effects, and that executive processes are predominately involved with exploration and not computation. This would explain why rising task difficulty in a motor study doesn't result in use of simplifying heuristics, as subjects continue to perform exhaustive computations (Snider et al., 2015). Complex processing strategies provide much better explanations of observed behavior in perceptual decision tasks as well (Shen and Ma, 2016).

Moreover, it was shown that differences in domain-specific performance in the equivalent decision tasks are an artifact of the different performance measures (Jarvstad et al., 2013). There are two common ways to assess rationality of an agent's behavior. If performance is measured by comparing agent's gains with those of a hypothetical "optimal" agent, as it is usually done in the perceptuo-motor decision literature, the differences appear to be rather small. For example, in (Jarvstad et al., 2013) the expected monetary outcome for the average participant is approximately 92% of the expected gain of an ideal agent. This almost optimal performance was observed for both perceptuo-motor and cognitive decision-making.

However, if one looks at the adherence to certain rationality axioms, as it is usually done in the economics literature, then the most likely conclusion is that humans seldom behave rationally. But this does not hold only for economic choice. When performed in a volatile environment, the presumably irrational context dependence was observed in both perceptual and motor decisions (Neyedli and Welsh, 2014; Summerfield and Tsetsos, 2015). A bias was also introduced with an asymmetric reward or cost structure (Hagura et al., 2017; Wu et al., 2015). Evidence speaks against inherently "better" performance in motor and perceptual tasks as opposed to classical decision-making. Even more so, domain invariant risk seeking behavior observed in children points to a possibly linked development of capacities for economical and visuomotor decision-making (Dekker and Nardini, 2016).

I want to draw attention to another, and a rather marginal observation in (Jarvstad et al., 2013); the difference in gains was driven almost exclusively by low-cost errors. That is, a suboptimal choice was more likely to occur when the difference between the values of alternative choices were smaller. This can be explained at a high level description. If the difference is pint-sized the loss incurred by a suboptimal choice might be too small to be worth the effort of a precise-enough computation. The observation is closely related to the observation mentioned above, where the size of an information set doesn't matter as much as how well does the information discriminate between the alternatives (Glöckner and Betsch, 2012).[14]

There's also a lower level explanation. Taking into the account neural noise corrupting internal representations, it comes as no surprise that the closer the representations, the more likely it is that the comparison will be swayed by it. It turns out, that contextual dependence in economic choice is well explained by *normalization*, a class of ubiquitous neural coding mechanisms first proposed to explain phenomena observed in a primary visual cortex (Carandini and Heeger, 2012). According to the Barlow's efficient coding hypothesis such mechanisms evolved from optimizing information processing under neurophysiological constraints (Carandini and Heeger, 2012; Louie et al., 2015).

[14] It is not clear whether the related model captures the link between discriminability and likelihood of an error, or if it just predicts longer reaction times.

> Barlow's efficient coding hypothesis proposed that sensory systems exploit widespread statistical regularities in the distribution of the sensory environment [...] regularity-induced redundancies in the incoming information stream are removed by sensory systems, increasing the independence of neural responses to different stimuli (thus maximizing information and increasing efficiency) (Louie et al., 2015).

We speak of cognitive, perceptual, and motor decisions separately to label the task relative to the main source of uncertainty. A question arises naturally if there is something common to how we process the probabilistic information from these different sources. Turns out there is! Neural correlates of probabilistic weights of risky prospects are represented in medial prefrontal cortex (mPFC), both when learned as a movement variance in a motor lottery, or given explicitly in a classical economic task Wu et al. (2011). Thus, it shouldn't be surprising to learn that the probabilities are represented in the same way, that is, as linear transforms of logarithmic probability odds, regardless whether subjects are performing a cognitive, perceptual, or a motor task (Zhang and Maloney, 2012a).

Secondly, an additional source of uncertainty was recognized in decision neuroscience literature, beside the nondeterminacy of the task environment, namely the fluctuations in neuronal signaling. Like perception and motor control, cognition is realized on the basis of neural computation. This is important, because uncertainty arises both due to a task-induced risk, and noise present in neural circuits. However, we're dealing with two different beasts here; while the former is often a "known unknown", the latter is always an "unknown unknown". By virtue of being realized in a neural circuit, any decision mechanism is bound to deal with a noise-induced ambiguity.

2.4 Neurons, circuits, and uncertainty

Some probabilities are external and relate to an uncertain structure of the environment. They can be either learned through experience, or communicated according to simple conventions. It was noted above that perceptual and motor decision-making come with additional stochasticity through variability in perceptual stimuli and motor execution.

There's an additional source of uncertainty that pertains to the cognitive decision-making as well. A certain amount of irreducible stochasticity is due to noise in neural representations of value. A brain doesn't compute with respective probabilities, but the mechanisms for value computation have evolved around mitigating such noise. The normalization models are conceptually appealing because they provide an account of how such computations could be robust and respect metabolic efficiency at the same time.

An action potential is the basic unit of neural signaling. It is a rapid change in electric potential, propagating along the neuronal membrane and possibly carries over to another cell. That happens only if at some point alongside neuron's membrane voltage between neuron's inside and outside gets sufficiently high to trigger it. After such an event occurs, it takes some time for neuron's membrane to reach its initial potential.[15] This has to happen before a new action potential can take place. If the voltage between two sides of a neuron's membrane doesn't get sufficiently low, a new rise in voltage cannot occur.

The time it takes a membrane to become excitable again limits the number of times a neuron can fire in a given period. This physiological constant provides an upper bound on neuron's signaling capacity. Consequently, the

[15] In fact the membrane's voltage potential needs to get even lower, *i.e.* the membrane needs to get "hyperpolarized".

absolute number of spikes cannot adequately represent wide-ranging quantities. Normalization is a simple way to extend the dynamic range of a neural signal. Since it was first proposed to explain a phenomena in a visual system (Carandini and Heeger, 2012), I'll use a retinal neuron as an example.

Sensitivity of a neuron responding to the light intensity in a certain small area of an animal's visual field is adjusted according to the activity of selected neurons responding to the close surroundings of the area. The greater light intensity is reported by these neurons, the more intense light will be needed to elicit the same response from the "normalized" neuron. Generally speaking, neuron's activity is normalized with respect to the summed activity of a broader neuronal pool. This allows the same neuron to signal small changes in light intensity across a wide wavelength spectrum. In contrast, a neuron that would invariantly map a specific wavelength to a specific response, could only exhibit a much more coarse-grained sensitivity.

We can easily convince ourselves that similar mechanisms come into play during a decision process if there are to be neurons dedicated to representing value. Throughout their histories decision-makers encounter options of varying worth, ranging from few cents to millions and above. If value were to be represented on an absolute scale, differences below certain magnitudes would be inevitably obscured due to physiological constraints of neural coding. Empirical evidence for involvement of such mechanisms in a decision process further relates sensory processing with economic choice. It also showcases the adaptiveness of neural decision-making machinery and provides a low-level mechanical explanation of contextual dependence of economic choice (Louie et al., 2015).

A simplified example is provided that nicely relates the previous discussion of discriminability with a claim that, as a rule, humans don't ignore attained information. Assume a primitive neurobiological system is choosing between items of differing worth. In its simple neural circuit, there is a value-neuron for each item representing the corresponding value. All value-neurons project to a choice-neuron, but also to each other. This allows them to normalize their response before they signal it to a choice neuron. Assume further, a choice-neuron communicates over channels corrupted by noise sampled from a distribution with zero mean and non-null variance. That is why value-neurons sometimes over- or under-sell the item they represent.

Presence of noise explains why a probability of the system choosing a less valuable option is higher, the harder it is to discriminate between options. Imagine that the system, when choosing between two options, consistently choses one option over the other. Offering it a third alternative would bring the normalized values of the first two items closer together and this could be enough for a break-down of the choice consistency.

Normalization was suggested (Tsetsos et al., 2016) as a plausible building block of a more elaborate neural algorithm called "selective integration" (SIA), which is built upon a concept of a sequential sampling model (SSM). A generic SSM assumes that information about choice values is sequentially sampled, from environment or memory, and then accumulated with previous samples until a decision threshold is reached. Values are sampled interchangeably with respect to a different option attribute. There could either be multiple value representations for each item, or a single quantity, representing the difference between accumulated across-attribute values of competing choices.

SIA simply states, that at each time values are sampled, a comparison occurs (*e.g.*, by a normalization mechanism) and the value of a local loser is discarded. That is, if two sample inputs are compared, such that $I_1 > I_2$, then at latter stages the system will integrate partial values I_1 and $\alpha \cdot I_2$, with $0 \leq \alpha \leq 1$. Since the value information is carried only by the magnitude of the respective accumulated quantity, this can be seen as effectively ignoring some information. At first it would seem that this is much in accordance to the assumptions behind heuristic models.

> A canonical argument states that such violations of decision theory [...] disclose fundamental limitations in human processing capacity and of the executive system (Tsetsos et al., 2016).

However, psychophysical experiments showed that this selective integrating is not due to processing limitations of human decision-makers as the bottleneck assumption would predict (Tsetsos et al., 2016). Simulations show that SIA actually increases the accuracy under the condition of late noise, that is, noise that is added to sampled values at the time of integrating them in the overall accumulated value.

> These findings suggest that violations of rational choice theory reflect adaptive computations that have evolved in response to irreducible noise during neural information processing (Tsetsos et al., 2016).

It is interesting to contrast SIA with a generic heuristic. Selective integration predicts that information is sampled sequentially and that some of it is eventually ignored. Both of these seem to support the conception of bounded agency currently in place. However, information is not ignored because of computational intractability or environmental ambiguity, but to guard against noise in neural circuits. Secondly, information is ignored only *after* it is processed; no computation is forgone like in a heuristic model. What remains to be shown is that sequential sampling, doesn't imply sequential processing, invoking again the distinction between information acquisition and processing.

Sequential sampling models are linear simplifications of the pooled mutual inhibition model (Bogacz et al., 2006). Details aside, the two distinctive features of such a circuit model are recurrent connections and mutual inhibition (Hunt and Hayden, 2017). Feedback connectivity is considered as the most likely architecture to implement normalization, especially because of abundant recurrent connections in cortical areas (Louie et al., 2014). If comparison of choices is indeed implemented as a competition via mutual inhibition (as predicted by the computational models), modulating the ratio between GABA and glutamate, the two neurotransmitters most commonly used by neurons for inhibition and excitation, should affect the decision process. Indeed, this was observed when the ratio was disturbed in ventro-medial prefrontal cortex (vmPFC), an area commonly implicated in value-based decisions (Strait et al., 2014; Wu et al., 2011).

> [V]alue-guided choice is governed by a competition by mutual inhibition that is mediated by a balance between GABAergic inhibition and glutamatergic excitation in the vmPFC (Jocham et al., 2012).

Increasing the ratio in favor of excitatory glutamate, subjects became worse at discriminating between choices with similar values. In other words, they became more susceptible to noise. This is the exact result (Hämmerer et al., 2016) first predicted with a detailed simulation of pyramidal neurons, and then showed experimentally with a transcranial direct current stimulation (tDCS) of vmPFC. Applying tDCS excites neurons, acting against the force of inhibitory signals. This indiscriminately increases baseline firing rates of neurons, making them more susceptible to background noise. Of course, mutual inhibition alone cannot explain this effect, because applying tDCS also excites the inhibitory neurons, thus the two effects should more or less cancel out. However, the effects are not symmetrical because of the strong recurrent connections of pyramidal neurons that amplifies the excitatory signal beyond the increment in an accompanying inhibitory signal. This experimental evidence thus speaks in favor of architectural assumptions of strong recurrent connections and ubiquity of competition through mutual inhibition.

Appealing to recurrency, we can now see why sequential input doesn't imply sequential processing. Neurons in recurrent networks not only receive input from other network units, but also receive their own previous output. This allows such networks "to show sustained memory for inputs long after they have been removed, allowing temporally extended computations to be performed on sequential inputs" (Hunt and Hayden, 2017). Neurons can "remember" their output by receiving it at a later time as an input from a recurrent pathway. Hence, neurons act both as mnemonic and processing units. This would mean that there is no dedicated architectural segment that would correspond to a processing bottleneck. Because mnemonic and processing units are not implemented separately, there is no need to shuffle data between the two, effectively doing away with a possible neural analog of a von Neumann bottleneck.

> The key insight [...] was that neuron-like units that performed biophysically plausible computations and were connected in simple ways could perform astonishingly rich computations. Such systems do not have dedicated memory and processing subsystems, unlike other computing architectures (Hunt and Hayden, 2017).

Collocation of memory and processing is also a distinct feature of the so-called memristive systems. That is, nonlinear dynamic systems that can be described as a generalization of a "memory resistor", or "memristor" for short. One such memristive system, curiously enough, was proven to be the Hodgkin-Huxley model of a neuron (Chua and Kang, 1976). It was only many years later that a memristor's physical existence was demonstrated (Strukov et al., 2008). Nowadays, many neuromorphic electronics[16], that is, electronic circuits and devices whose architecture relies heavily on the known principles of neurobiological organization, utilize memristors to build *non* von Neumann like computers (Schuman et al., 2017).

> However, the inevitable comparisons of this [von Neumann] architecture to the human brain highlight significant differences in the organizational structure, power requirements, and processing capabilities between the two. (Schuman et al., 2017)

Simple neural mechanisms have evolved on the basis of architectural principles that allow high, yet metabolically efficient processing capacity. The distinction between digital computers and neural systems is based on the differences in basic computational units and their organization. Importantly, these differences lead to different processing limitations. Ignoring information by forgoing computation would be beneficial if executed on a sequential and modular machine. Neural architecture is more likely organized around principles of high parallelism, recurrent connectivity, and mutual inhibition (Hunt and Hayden, 2017). These suggests computations based on pooling, competition through mutual inhibition, and normalization. These in turn provide a good low-level explanation of observed contextual effects and cognitive costs of ignoring acquired information, and suggest evolution towards robust decision-making under circumstances of unquantifiable uncertainty.

2.5 Conclusion

The widely accepted view in the heuristics literature states that humans tend to make decisions following simple algorithms, the so-called heuristics. Algorithms

[16] See Chapter 4 for further discussion of NE.

are described as simple if they ignore part of the available information. There are two purported benefits of doing so. Namely, minimizing the amount of computation, and achieving robustness in face of unquantifiable uncertainty. I argued that the latter is a novel theoretical contribution of the FFH research program, while the former is a legacy of Simon and Newell's work on classical AI. In line with the notion of ecological rationality its proponents maintain that the two facets of simplicity relate to two types of constraints imposed on an agent, inherent and environmental. In their view, minimizing computation is a consequence of limited cognitive resources, whereas robust algorithms are an answer to epistemically unfriendly environments.

Relying on results from both psychological and neuroscientific literature I argued that this view of bounded agency needs to be revised. In a nutshell, it seems that there is little reason to believe that simple heuristics that ignore information minimize resource consumption, as they're related to additional cognitive effort and neural activity. While the neuroscientific literature corroborates the claim of evolved trait of robustness, it emphasizes the neural noise rather than environmental ambiguity as the source of irreducible uncertainty. Moreover, there seems to be some parallels between the idea of ecological rationality from the heuristics literature, and Barlow's efficient coding hypothesis.

This gives further support to the intuition that decision processes have evolved to exploit the statistical regularities of the environment. More importantly, it could also suggest that decision algorithms are such that the amount of information is boundedly optimized relative to the metabolic cost needed to encode it. This relates closely to the question of how costly is it exactly to ignore already acquired information. Is ignoring a cue less costly than ignoring a more predictive cue? If so, what if the more predictive cue is strongly correlated with a third cue, whereas the less predictive cue is not? How does the causal knowledge of the environment affect these results? What happens when it is the combinations of multiple cue values that are truly predictive of the environmental state? What if a cue is introduced with a certain predictive strength, but then the strength changes?

Ignoring a cue that is highly correlated with another one is discarding a lesser amount of information as when ignoring that same cue in the absence of the other. That's why I would expect that ignoring a less predictive cue is more costly, as soon as there are two or more cues that are strongly predictive on their own, as this entails their correlation. Moreover, it would seem that these effects would be stronger in environments in which cue relationships would be determined by an "expected" causal structure of the environment and weaker when they would contradict it.

I've also argued that the current measure of simplicity based on algorithmic complexity analysis assumes wrong computational architecture. Namely, the heuristics are analyzed as if they were run on a sequential machine similar to a digital computer, in line with core assumptions of the classical AI research. I believe that the use of ACT-R (R. Anderson, 2007; Fechner et al., 2018) and similar cognitive architectures to evaluate the costs of using a heuristic is a step in the right direction. However, it's been long since the rule-based models had their own corner on the market. The alternatives include network (Glöckner et al., 2014) and circuit models Genewein and Braun (2016). For such models circuit complexity analysis would probably be more suitable (Wegener, 1987). Particular attention should be given to work on the complexity of neural networks (Parberry, 1994). To the best of my knowledge, this is the only rigorous attempt at understanding of how does complexity of neural networks compare to complexity of conventional computers.

3 On analog neural computation

I this chapter I discuss a recent account of analog neural computation and its underlying theory of analog and digital representations. I raise several concerns regarding the empirical validity of a premise used to argue that neural computation is analog. Furthermore, I present some conceptual tensions between this account and typical computational explanations found in neuroscience. I trace these issues back to the underlying definitions of representations.

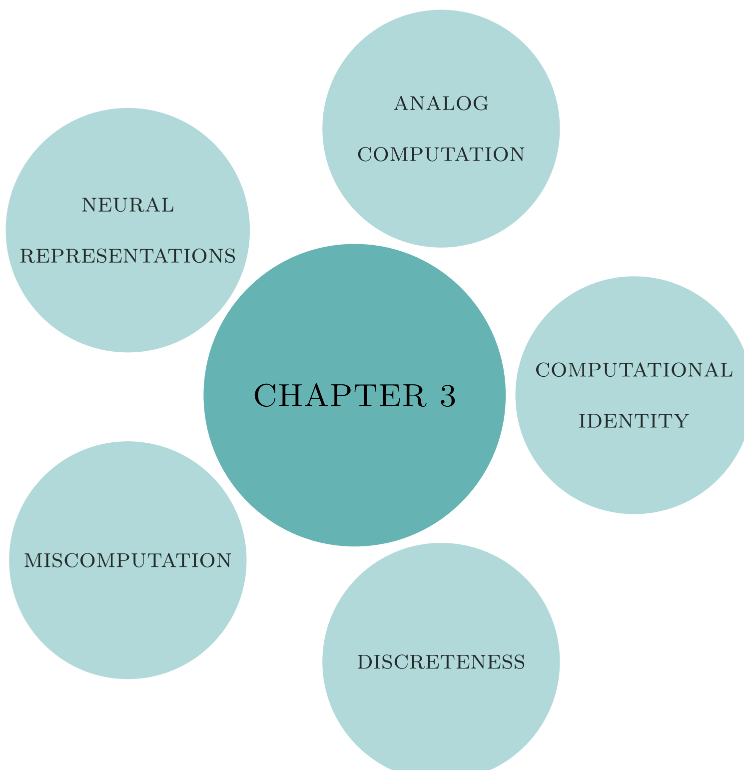

Figure 5: Graphical summary of Chapter 3.

3.1 Introduction

In a recent issue of *Minds and Machines*, Maley (2018a) argues that we should ground computationalism about brains in analog computation, which he defines as "the manipulation of analog representations". The argument goes as follows:

(M1) Analog computation is manipulation of analog representations.

(M2) Analog representations can be continuous or discrete.

(M3) Digital representations are discrete.

(M4) Brains compute with continuous signals.

$\therefore$ Brains are analog computers.[17]

$\therefore$ Brains manipulate analog representations.

Premise (M4) is an empirical statement and will be the focus of the next section. I will examine in more detail the reasons given for its acceptance and argue against its empirical validity as well as conceptual fruitfulness. The two central concerns are that Maley's approach is not able to recover a notion of miscomputation[18] and postulates too strong of a notion of computational identity[19]. More specifically, Maley's position rules out the existence of neural

[17] There is a hidden premise that there are only two types of computation – namely digital and analog. In fact, Maley advertises the approach as an alternative to postulating a third type of computation (neither analog nor digital) that would include neural computation (see Piccinini and Bahar, 2013).

[18] If a computer is computing a function f for an input i and the output $o \neq f(i)$, then one might want to say that the computation has failed. In other words, a "miscomputation" has occurred. See also (Piccinini, 2015, p. 12-14)

[19] The notion of computational identity is tied directly with the notion of miscomputation. Imagine a computer that successfully computes f for a given input at some time, but for

noise without offering a new conceptual device that could fill its role in typical neuroscientific explanations.

The reasoning is slowly laid out in more detail throughout the chapter. However, to assist the reader I will shortly discuss the crux of the argument here. We should begin by asking ourselves what does it mean for a computer to be continuous (in a sense that excludes digital computers). It should be noted that ordinary digital computers can perform some *precise* computations with continuous (and even transcendental) numbers. Insisting that brains are continuous computers and digital computers are discrete, must thus be meant in a sense that the variables denoting brains' computational states are continuous.

I spend a significant portion of the text arguing that physical states of brains are not continuous, that is, that there are only countably many possible physical brain states. This in itself would invalidate the premise of continuous computational states – cardinality of the set of computational states cannot be greater than the cardinality of the set of the underlying physical states.[20] However, one does not even need to go that far. The point is that not all fluctuations matter – even if the physical quantity is changing continuously, not all of its changes carry information.

This is where the noise comes in or so I argue later in the text. If we agree that there are random fluctuations (noise) that can be present but do not

some reason miscomputes when given equal input at a later time. This two events are two attempts, one successful and the other failed, of the computer performing *the same* or "identical" computation.

[20] I agree that it is often more practical to talk about brains as continuous systems, but this is not enough for a categorical difference between brains and digital computers. Moreover, it hardly even holds as a "practical difference", given that digital computers are also being increasingly more often described using analysis/calculus rather than discrete math – *e.g.*, when we talk about parasitic capacitances, quantum tunnelling, *etc.*

affect the computational state, then we cannot say that there are uncountably many computational states. That is, if the state is determined by a certain value of the membrane potential $\pm\epsilon$, then as long as $\epsilon > 0$, the respective variable describing the computational state is discretized.

The point about miscomputation is that sometimes we want to say that the same computation was performed in different ways (some successful, some failed). However, if we insist that every infinitesimal change constitutes a different computation, then every computation will be mapped to an exact physical state.[21] And we'd also be forced to agree that every computation is successful if performed.[22] On the contrary, I'd like to say that performing the same movement multiple times amounts to performing the same computation multiple times and that the variation in the outcome is the result of the noise on top of that computation. In this sense a notion of noise is integral to neuroscientific explanations involving miscomputation, and a theory postulating that brains compute with continuous signals will likely fail to make sense of its use.

Premises (M2) and (M3) follow from the definitions of analog and digital representations given in (Maley, 2011). These, together with the premise (M1) will be the focus of the third section. There I will first go through the definitions of the two types of representation in more detail. I will then argue on examples that the definitions are ill-founded and are in particular not applicable to a wide class of neuronal representations. Most importantly, by further explicating Maley's definitions of representations, I will show how the issues with miscomputation and computational identity are more general and might have something to do with defining a type of computation as a manipulation of the respective type of representations, as exemplified by (M1). Finally, Maley

[21] Or an exact sequence of states, or rather an evolution of a physical quantity in time.

[22] Unless one would somehow argue that the same physical evolution can constitute both successful and unsuccessful computation.

(2011, 2018a) offers a handful of useful notions related to a notion of representation that lead me to explore interesting connections to other open questions in philosophy of computation and neuroscience, and their intersection. An outlook for future research is given in the conclusion.

3.2 On continuous neural computation

My concerns regarding (M4) can be stated very generally, namely that neuronal processes operate on discrete intervals and with finite amounts of resources.[23] The communication and computation in neural systems are realized by changes in electro-chemical gradients, by means of ionic current, neurotransmitters, and protein modulation which effects the electrical properties of neurons or acts upon the system in some more general way.

One need not even argue that the neural computation ought to be described as finite. The premise (M4) is already falsified if the computation turns out to be discrete. With that in mind, I'd like to point out the charge quantization principle which simply states that charge is quantized and therefore any amount of charge will be countable. This holds even more so for the case of neural systems, since the charge is carried by an ionic current. The amount of ions is surely a discrete quantity, and most likely finite as well. This holds for the neurotransmitters, G-proteins and other cascading chemical components that play a role in neural computation. All these physical quantities surely come in abundance and their role in neural function, broadly construed, is well established (see Bear, 2016).

Moreover, neural computation is not free. It requires energy, for example in form of ATP (adenosine triphospate) consumed by ion pumps – neural

[23] See also (Piccinini and Bahar, 2013).

mechanisms that are vital for repeatedly establishing an electro-chemical gradient between the outside and inside of a neural membrane that is crucial for propagation of *any kind* of electrical signal. Given that there will be only a finite number of these pumps working under specific energy, space, and time constraints, it is hard to see how electro-chemical gradients could change continuously.

Even then, the mere presence of a continuously graded quantity by itself is not necessarily computationally relevant. This is related to the fact that changes of a neuron's physiological state, at least on the relevant time-scales, take time. That is, they're not instantaneous.[24] Hence a neuron can only assume a finite number of states in a finite amount of time, even if acted upon by a non-discrete quantity. This should be kept in mind when assessing the reasons given in support of (M4):

> [T]here are several ways in which the continuity of a phenomenon related to neural spiking plays an important role in neural signaling. First, unlike digital computers, there is no discrete 'clock' that determines when a physical change in a circuit element should count as a change in signal. Second, unlike digital computers, the rate at which physical changes in a circuit element occur–which itself is a continuous quantity–can have effects on downstream elements in the system. And third, unlike digital computers, the precise shape of the voltage waveform of a circuit element going from 'off' to 'on' can have differential effects on other elements in the system. (Maley, 2018a, p. 85)

[24] For example, it takes time for an ion pump to open or close, and a neurotransmitter molecule or an ionized particle to travel a certain distance (including "quantum jumps" Minev et al. (2019)).

Three claims can be identified in the above passage, and I will go through them one by one. For each premise I will present empirical results that undermine the validity of inferring (M4). Moreover, I will present conceptual arguments against characterizations of both neural and digital computations as implied by the premises (M4.1-3):

(M4.1) "[T]here is no discrete 'clock' that determines when a physical change in a circuit element should count as a change in signal." (Maley, 2018a)

(M4.2) Temporal and rate neural codes use continuous variables.

(M4.3) Continuous variation of membrane voltage waveforms is computationally relevant.

At least superficially, the well-researched notion of neural oscillations seems to be at odds with (M4.1). On the other hand, a direct comparison would be a bit underhanded. A clock in a conventional digital computer has a much more determinate effect on the efficacy of a signal than neural oscillations do in brains. That is, in a computer a signal either arrives at an appropriate time and is dealt with, or it doesn't and is simply ignored. The same could not be said for the brain. Nonetheless, one could argue that this is a difference in degree and not in kind and moreover, that both mechanisms seem to be in place to perform the same function – synchronization. Similar to conventional computers, periodic changes in membrane voltage potential following a frequency shared among multiple neurons allow temporal coordination between larger groups of neurons.[25]

The oscillation-related fluctuation of the membrane potentials in

[25] Other type of oscillation occurs when an assembly of neurons fire periodically and in synchrony.

> the participating neurons continuously and predictably biases the open-time probability of a multitude of voltage-gated channels. This design is an energy-efficient solution for periodically elevating the membrane potential close to threshold, providing discrete windows of opportunities for the neuron to respond. If the input is not appropriately timed, however, it is ignored altogether or the response is delayed. (Buzsáki and Draguhn, 2004, p. 1928)

Neural oscillations have been commonly researched in the hippocampus. Focusing on the latter, there are particularly telling results that speak against (M4). In particular it would appear that only finitely-grained time differences are computationally relevant for neural computation. This follows from another similarity between *functional* importance of computer clocks and somewhat incidental *effects* of physiological limitations.[26] There are reasons to believe the "discrete windows of opportunities" in the hippocampus are determined by a physiological constant of about 10–30ms (Harris et al., 2003) that matches:

1. the membrane time constant of pyramidal neurons (Spruston and Johnston, 1992);
2. the period of the hippocampal gamma oscillation (Chrobak and Buzsáki, 1998; Csicsvari et al., 2003);
3. and the time scales of molecular mechanisms underlying long-term potentiation and depression[27] (Magee and Johnston, 1997).

[26] That is, "another similarity" that speaks in favor of the "difference in degree, but not in kind" argument.

[27] LTP and LTD are two related molecular mechanisms underlying learning through modulation of synapse strength also known as "spike timing-dependent plasticity" (STDP). Put

In short, neural oscillations are ubiquitous brain phenomena, not limited to the hippocampus (Wang, 2010; Feurra et al., 2011; Polanía et al., 2014b) and the related research suggests that the "clocking mechanisms" in brains are likely to reflect some physiological limits and are thus not arbitrary.[28] Moreover, these "biological reasons" are analogous to why clocks are used in digital computers. Due to physical properties of computing mechanisms the signal has to be kept constant for a certain period of time, and there has to be a certain time difference between two different signals in order to achieve stable computation (*e.g.*, by avoiding race conditions).

It is important to note that using a clock is *only one* possible solution. There's nothing to digital computers that would make the presence of a clocking mechanism necessary. Research on asynchronous digital computers is an active field, and as old as its synchronous counterpart (Nowick and Singh, 2015). The absence of a "discrete clock" is not enough to conclude that a computing mechanism under inspection is not a digital computer. Moving forward to (M4.2) I consider two possibilities of how continuous variables might figure in neural

simply, if a presynaptic neuron fires shortly before a postsynaptic neuron, the strength of the synapse increases, but if it fires shortly after, the synaptic strength decreases (Bear, 2016, p. 874-9). The "coincidence detection" usually involves binding of glutamate to NMDA channels and the strength of modulation is proportional to the amount of Ca^{2+} ions entering the cell through the channel. Glutamate binds to NMDA receptors for many tens of milliseconds, likely surpassing the amount of time the channel is actually open. Thus the difference in arrival of two spikes before the depolarization of the postsynaptic membrane, will only count if they occur sufficiently long apart and early enough for there to be difference in related durations of glutamate being bound while the membrane is depolarized. More generally, different arrival times will only have different effects if spikes arrive far enough apart for there to be a sufficient difference in the amount of Ca^{2+} passed through the channel. Thus the lower bound of temporal precision corresponds to the time it takes for a minimal effective amount of Ca^{2+} to pass through a NMDA channel.

[28] This will prove relevant when discussing (M4.2).

computation *via* temporal and rate codings.

> Two separate, well-known phenomena illustrate the importance of continuity, rather than discreteness, in neural signaling. The first, called temporal coding, occurs when the time between individual occurrences of APs has some functional significance. The second, called rate coding, occurs when the overall number of APs within a given duration–the frequency of APs–has some functional significance (Maley, 2018a, p. 83).

Given any time interval a number of spikes that occur will always be an integer – a spike either occurs or it doesn't. Moreover, as discussed above, the rate interval is likely bounded by certain physiological constants. Thus variation in the length of the interval which could be the source of variation in the rate code is unlikely to really provide a continuous spectrum. That's why the lack of evidence of continuously varying rate code comes hardly as a surprise.

The idea behind the notion of temporal coding is that neural code resides in temporal differences between different spikes. That is, that information is encoded in the amount of time that passes between arrival of two given spikes, rather than (or in addition to) encoding it with the number of spikes received within a certain time interval, *i.e.*, rate code. Assuming that time is continuous, then so should be spike timing differences. An additional argument is wanting, namely that the temporal coding is absolute. However, it seems that temporal coding is more likely to be relative or based on the order, rather than on "precise timing" of the spikes' arrivals (Stiefel et al., 2012; Butts et al., 2007). In other words, temporal coding has more likely to do with discrete combinations, rather than continuously precise temporal differences. Moreover, the term "precise spike timing" in neuroscientific literature discussing temporal coding is used in a rather peculiar way. Namely, "precise" usualy denotes temporal resolution on a millisecond scale (Panzeri et al., 2010).[29] Finally, it is reasonable to expect that the "coincidence" detection mechanisms operate on a millisecond scale (possibly lower, but definitely finite, see footnote 27). Therefore, smaller, let alone "infinitesimal", timing differences are unlikely to be relevant.[30] This point is worth reiterating in order to avoid possible confusion. I do not argue that inter-spike intervals or time in general are not continuous, rather I claim that the relevant time differences are discrete insofar as neurons are not sensitive to infinitesimal temporal changes.

There are also more conceptual reasons to reject (M4.2). It should be noted that any kind of code will need to be decoded sometime along the downstream signaling path. In order for an alleged continuity in neural representations to play a role in neural computations, the decoding process needs to preserve the continuous nature of the represented variable. Put simply, conti-

[29] To the best of my knowledge, the smallest time resolution reported is on the scale of microseconds (see Bale et al., 2015).

[30] To illustrate the point, consider polychrony as an example model of how neural temporal code could be used to encode information. The term describes "time-locked but not synchronous" (Izhikevich, 2006, p. 245) propagation of spiking activity through a neural network. Overlapping subsets of neurons are organized into "polychronous groups" (PG) which produce a "regularly repeating chain of activity" (Eguchi et al., 2018, p. 546). The idea is that the number of PG greatly exceeds the number of neurons (or even synapses), yielding greater representational capacity. Importantly, the formation of groups occurs under constraints imposed by differences in axonal conduction delays of different neurons projecting to the same target. It was proposed that STDP "can select matching conduction delays and spontaneously organize neurons into such groups" (Izhikevich, 2006, p. 249). Due to the nature of mechanisms behind STDP the time differences between delays under a certain bound simply won't be recognized by the system and are thus not relevant (see footnote 27). Unsurprisingly, the authors report millisecond temporal precision in a network with an order of magnitude lower scale of axonal delays. Finally, it is worth noting that discussed time-differences are still referred to as "precise", despite the apparent coarse-grained millisecond precision (Izhikevich, 2006; Eguchi et al., 2018).

nuity needs to be found in the temporal and/or spatial dendritic summation at a "decoding" neuron. For this reason (M4.2) stands and falls with (M4.3) which states that the continuous variations in a neural membrane potential are computationally relevant.

There are two ways to argue against (M4.3) – either by denying that neural membrane potentials vary continuously, or by rejecting the idea that *all* variations play a role in neural computation. An argument of the first type can be made for a large population of neurons that communicate through chemical synapses:

> [Excitatory postsynaptic potentials] at a given synapse are *quantized*; they are multiples of an indivisible unit, the *quantum*, which reflects the number of transmitter molecules in a single synaptic vesicle and the number of postsynaptic receptors available at the synapse. (Bear, 2016, p. 133)

While Maley (2018a) does not explicitly claim that chemical neural communication constitutes continuous computing, an argument is lacking as to why (M4) should qualify generally for brains as a whole. Perhaps we should entertain a more modest claim, that part of a neural computation is based on manipulation of continuous variables. Namely, computation realized by the so-called "non-spiking neurons" and electrical synapses.

> Not all inter-neural communication is based on action potentials. Some rely on direct electrical signaling across synapses via what are called gap junctions. This signaling is not all-or-nothing, but continuous. (Maley, 2018a, p. 80)

While this is arguably possible, albeit very unlikely, I'll argue that any account of neural computation that characterizes it as continuous computation, is theoretically unfruitful and leads to strange commitments.[31] Let's assume for the sake of the argument, that membrane voltage waveforms do in fact vary continuously. From (M4) it follows that infinitesimal changes are computationally relevant.[32] Thus a following argument can be made:

(B1) Precise membrane voltage waveform is computationally relevant.

(B2) Membrane voltage is strongly dependent on the ratio between inside and outside concentrations of K^+.

(B3) Eating a banana increases blood-levels of K^+.

∴ Banana intake is *always* computationally relevant for motor control.

One might argue that dietary choices would be less likely to impact computations confined to brains, due to the brain-blood barrier, and potassium spatial buffering and other astrocytic mechanisms for regulation of the potassium gradient. However, a similar argument could be constructed by replacing (B2) with:

(B2*) Membrane voltage is strongly dependent on body temperature.[33]

[31] On that note it is important to emphasize that an argument applies more generally to any kind of representational medium, without limiting the argument to neural membrane voltage waveforms.

[32] Arguably, voltage levels in digital computers also vary continuously. What makes them discrete is the fact that only a countable number of states are computationally relevant.

[33] See Goldman-Hodgkin-Katz voltage equation and Nernst equation.

All that's left is to just fill in for (B3*) something that influences body temperature and a similarly counter-intuitive conclusion will follow. Before discussing what exactly makes such conclusions problematic, let's have a look at a related consequence of adopting (B1) which is directly implied by (M4). Consider a widely adopted idealization that neurons are well described as Gaussian channels, to which Maley (2018b) seems to subscribe as well:

$$R(t) = N(t) + Z\,; \quad Z \propto \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

It simply states that an output of a neuron at certain time will be equal to the neuron's "actual response" plus some noise sampled from a zero-mean normal distribution. The noise distribution is irrelevant past its ubiquitousness in neuroscientific literature. However, even adopting a weaker notion of noise without assuming any distribution will still turn out to be meaningless, if one is to accept (B1). If the precise shape of the neuronal membrane voltage is computationally relevant, it is hard to make sense of what neuronal noise would be. Conversely, if we accept that some intervals proportional to noise (rather than point-values) are mapped to computational states, (M4) must be rejected. In other words, any degree of robustness to noise will lead to binning, that is, discretization of the related computational variable.[34]

[34] A purported counterexample to this argument are the neuronal instances of different frequency filters. Removing a component of a continuous signal corresponding to a certain frequency band will not necessary result in a discrete quantity. However, this counterargument assumes that the output of the filter is itself noiseless or deterministic. Existence of any such neural filters is dubious. By saying that such a quantity is inevitably discretized I simply mean that for each possible output of the filter μ, the amount of the input quantity could be either increased or decreased, say by ϵ, such that the μ would not change. In this sense the measurement is discretized to the intervals of length 2ϵ.

Sure enough, one might just bite the bullet and dismiss the notion of noise altogether. Since I granted the possibility of continuously changing physical states of a neural membrane, this becomes a dilemma about the definition of neural computation, rather than an empirical question. I take the general consensus to be that definitions should be judged by their usefulness. The main reason why a definition of neural computation based on (M4) doesn't meet my expectations has to do with an apparent inability to compensate for the conceptual convenience of postulating a noise component.

For one, I would like to keep the notion of noise because it allows us to properly explain miscomputation. A template explanation would be: "Neural circuit failed to compute a correct value, because there was too much noise." A good example of such explanations can be found in decision neuroscience. If a noise component is assumed, it allows explaining contextually dependent decision-making – and the so-called "preference reversals" – as a result of normalization, a "canonical neural computation" (Louie et al., 2013).

If a computational variable is truly continuous, then an infinitesimal change will already be computationally relevant and will result in a different computation. Accordingly, any computation is either performed precisely, or it is not performed. I struggle to see we could make sense of the notion of miscomputation under this understanding of neural computation. It seems that an intended computation, if performed, will never miscompute. It remains to be seen how one could characterize miscomputation when following (M4) for the cost of discarding the notion of noise. Relatedly, it would seem that the notion of computational identity becomes too strong. Allowing the possibility of miscomputation, an identical computation must be recognized in at least two exact neuronal states. That is, one exact state that performs a computation sucessfully, and at least one other that performs a computation unsucessfully.

To illustrate, consider a perceptuo-motor system that is attempting to realize a certain trajectory of a controlled plant, *e.g.*, a hand-reaching movement. The variance in neural activity behind system's multiple attempts to do so, could be characterized as motor noise. This is valuable, because in a failed attempt it is meaningful to say that the system tried to perform *the same* computation as during a successful attempt, but mis-computed.

3.3 Analog and digital representations

The account of analog neural computation given in Maley (2018a) defines analog and digital computations as manipulations of respective representations. The latter form a dichotomy that is described in Maley (2011). The resulting dichotomy is advertised as a useful alternative to the accounts that place the difference between digital and analog computation in the difference between continuous and discrete computation. After presenting the definitions, I will describe some general problems in addition to the empirical inadequacy of using the definitions in explanations involving neural computation. Interestingly, the problems of miscomputation and computational identity turn out to be more general and not confined to the considerations of applying the theory to neural computation. Analog representations are defined as follows:

> A representation R of a number Q is analog if and only if:
>
> 1. there is some property P of R (the representational medium) such that the quantity or amount of P determines Q;
> 2. and as Q increases (or decreases) by an amount d, P increases (or decreases) as a linear function of $Q + d$ (or $Q - d$) (Maley, 2011, p. 123).

As already noted by (Maley, 2011), the linearity might be too strong and could be replaced with monotonicity. I will assume this weaker requirement, since it seems necessary to accommodate examples of neural activity representing a quantity on a logarithmic scale (Gold and Shadlen, 2001; Zhang and Maloney, 2012b). Importantly, this definition allows discrete analog representations. In contrast to analog, a digital representation is defined as a tuple of:

1. a series of digits, each of which is a numeral in a specific place within the series; and
2. a base, which determines the value of each digit as a function of its place, as well as the number of possible numerals that can be used for each digit (Maley, 2011, p. 124-5).

A simple example should go a long way at illustrating the intended difference. Say a change in temperature is represented by a change in the size of a heap of pebbles. Every time a temperature drops by a kelvin, we take away a pebble, and every time it increases by a kelvin we add one. Representing is done by a quantity. Imagine now that the temperature is represented by a string of 0's and 1's. Adding a digit doesn't necessary change the represented magnitude. In fact, adding or removing either of the two digits, *ceteris paribus*, can result in a representation of a smaller, higher, or equal magnitude. Similarly, one can think about the length of the string - knowing only that the string is longer, shorter, or of unchanged length, leaves us completely ignorant to the changes in representation. Representing is done by a quality.

A digital representation is built out of elements of a specific set of structured quantities, where the structure allows us to distinguish between different members of the set, *i.e.* symbols – in this case interpreted as digits, organized as described in (Maley, 2011). Analog representation is not structured in this sense. That does not mean that whatever is used for a digital or an analog representation doesn't possess a quantity or qualities, respectively. Symbols are physical patterns, structured, rather than shapeless quantities, but quantities nonetheless. We can think of a myriad of ways we could organize our little heap of pebbles into strings of symbols representing a binary number, just by adding a bit of a structure. The converse is also true. A set of symbols could be used to form an analog representation. In that case, the structure of a string of symbols would play no role – only the amount of symbols would be relevant for the representation.[35]

This becomes particularly obvious when one considers a unary number system – in Maley's terms, a digital representation using number 1 as the base. In fact, unary numeral systems have been implied in certain neural systems, like song-production pathways in bird brains (Hahnloser et al., 2002; Fiete et al., 2004). But a unary digital representation is indistinguishable from an analog representation using the number of digits as a representing quantity. Should we say that in such cases, manipulating the representation constitutes both analog and digital computation? Such an outcome seems undesirable.

On a more favorable reading, the unary number system is not really a problem for Maley's classification. There are are two strategies one might adopt: 1) strengthen the definitions to obtain a more strict dichotomy (*e.g.* by requiring that the base is an integer[36] equal to 2 or greater); 2) accept fuzzy boundaries between the two notions and limit the problematic claims to the

[35] Consider an example of strings representing hexadecimal numbers. The two strings "0x00f" and "0x0000f" would be interpreted as representing the same value when used as digital representations, but possibly different values when only treating them as quantities, that is, as analog representations. Similarly, "0x00f" and "0x0f0" would be different digital, but equal analog representations.

[36] I assumed only an integer can serve as a base, given that in Maley's definition, base determines the number of possible symbols.

"non-degenerate" cases.

The second solution is not desirable and leaves us empty-handed. Maley (2011) advertises the definitions as particularly valuable for use in cognitive science. However, we've seen that a significant part of neural systems would fall under the class of "degenerate" cases. Wouldn't this defeat the very purpose of providing a definition of analog representation that could be used to talk about neural computation?

The first solution is unprincipled. While, for example, a binary system might be favorable over a unary representation in most applications, such choices reflect convenience (for example, when there is a need to represent negative numbers), and not a difference in kind. A separate argument is needed why the number 1 shouldn't be considered as a base number. Otherwise, we're running the risk of the definition arbitrarily excluding objects that intuitively qualify for the label.

This brings me to my second complaint about Maley's definitions of digital and analog representations. Restricting digital representation to place-value notation exclude cases of devices that intuitively should qualify as digital computers. Following Maley's definitions, digital computers encompass only devices performing numerical computation, excluding the cases of theorem proving, SAT checking, computer algebra, *etc.* I'll return to these examples below, when discussing computation with continuous variables.

Turning to the definition of analog representation, it is hard to see how it applies to neural representations. Maley needs to argue that the latter must not be representations of their own kind. Otherwise, neural activity would constitute a distinct kind of computation, as proposed by Piccinini and Bahar (2013). As things stand, this proves to be very hard for a very simple reason. A great number of neurons act as band-pass or band-stop filters (Hutcheon and

Yarom, 2000), essentially responding non-monotonically. The role of monotonicity was not made explicit by Maley (2011, 2018a), however, it is clear that the argument fails as long as one stands by the requirement.

There are two other important characteristics of representations as defined by Maley (2011) – representational format and representational resolution. A quantity used to represent is a "representational medium" (RM) and can be either continuous or discrete, depending on whether its physical states are countable or not. Hence, specifying a RM answers the question of what is *representing*. The "representational format" (RF) describes what is *represented* and it too, can be either continuous or discrete.

> For example, if I have several tons of sand for representing a number between 0 and 1, it might be most expedient to consider the representational format to be continuous, ranging over all real numbers between 0 and 1 (i.e. $Q \in \mathbb{R}$, $0 \leq Q \leq 1$). On the other hand, if I'm representing that same range of numbers with only a few hundred grains of sand (decreasing the available "resolution"), it might be better to consider the format to be discrete [...] Thus, the representational format may be continuous (in the case of using real numbers) or discrete (in the case of using hundredths), although the representational medium (grains of sand) is discrete (Maley, 2011, p. 118).

As evident from Maley's example, RF and RM are independent. It's almost trivial to state that a continuous medium can be used to represent a discrete format, but the case of a discrete medium being used for a continuous format is very telling. The question is whether digital computers can compute with continuous variables and the answer is a simple yes. While floating point arithmetics performed by an ordinary desktop computer shouldn't be controversial for Maley, two more examples will help drive the point home. First are the computable transcendental numbers already discussed by Turing (1936). Even more interesting is the case of symbolic computation, that is, computation using sign-value notation.

Digital computers often use sign-value notation to represent certain (mathematical) objects, for example, π. A sequence of digits with a determined sign-value representing π corresponds to an exact (transcendental) number that denotes a ratio of a circle's circumference u to its radius r, as expressed by the formula $u = 2\pi r$. Using such signs, instead of place-valued approximations, digital computers can perform *exact* computations with continuous variables for example to evaluate $\sin(\pi/4) = \sqrt{2}/2$, where the result itself is again expressed using a sign-valued representation. What does this tell us about (M4)? It seems that it has to be a statement about RM, since digital and analog computation are indistinguishable when only RF is known. More precisely, since physical states of a digital computer are (arguably) continuously graded as well, (M4) states that uncountably many physical states are actually used for computation – an unlikely state of affairs.

The last issue I'd like to discuss is related to the asymmetry between RF and RM. That is, the issue of the "resolution" of a representation. Resolution is effectively determined by the ratio of different representational states to the different states of what is being represented. For example, when representing 8 different digits, two pebbles will have a lower resolution than three pebbles. And three pebbles used digitally (used with a certain structure), will have a higher resolution than three pebbles used analogically.[37]

[37] This should hardly come as a surprise, given that analog representations are basically digital representations with base 1. Hence, as soon as there are $n > 1$ different RM states available, one can use the RM for a digital representation with base k, such that $1 < k \leq n$, with a guaranteed equal or higher resolution.

Imagine that you are given a contraption running an analog computation in the sense of (M1). The function being computed is such that the resolution of the representation (weight of pebbles) can make a difference with respect to the outcome of the computation. For example, depending on the RM used – either big pebbles that weigh 2g, or small pebbles weighing 1g – one might get different outputs for an identical input. Since, *ceteris paribus*, different resolution implies different representation, it follows that the same computation can be performed using different representations. Now, given that analog representation can be seen as just a digital representation with a minimal resolution – this leads us to conclude that a difference in the type of representation does not entail a difference in the type of computation. This brings us back full circle to a more general discussion about miscomputation and computational identity.

These intuitions become clearer, and hopefully less controversial, when considering an example with a digital computer. Assume I'm using an 8 bit unsigned integer representation and try to add 0xff and 0x01. Because of the overflow I will end up with a wrong number, 0x00. However, using the same RF(!) only with a greater resolution, say 12 bits, I will get the correct value, 0x100. Intuitively, *the same* computation was performed in both cases. Furthermore, a miscomputation occurred when the 8-bit representation was used and its occurrence is properly explained by the resolution of the representation being too low.

Arguably, a similar situation arises when using analog-to-digital (ADC) or digital-to-analog converters (DAC). In either case we have a type of representation that is a product of a computation of another type. For example, a digital computer coupled with a DAC will result in an analog representation that was computed digitally. The converting processes are especially relevant for the topic of neural computation, as some authors have argued that neurons perform both types of conversion (Sarpeshkar, 1998). These considerations lend themselves naturally to a view that neural computation shouldn't be defined over what is *representing* or how it *represents*, but rather over what is being *represented* (Egan, 2010).

Nonetheless, one might argue that this is an unrelated issue. After all, I am talking about types of computation, and not types of computers. The (M1) could be changed accordingly:

(M1*) Analog computers manipulate analog representations.

While a fair point in itself, the issues I've outlined above appear to haunt many practical applications of the theory. For example, when arguing whether brains are either analog or digital computers. The problems with proper specification of a type of computation point to a possible underdeterminacy of leveraging observed computation to infer the type of the computer performing it. Hence, if I am not able to distinguish between analog and digital computation (especially in the "degenerate cases" that are ubiquitous in neuroscience), I won't be able to determine the type of computer based solely on the observations of performed computations.

3.4 Conclusion

The arguments presented in this chapter might be considered somewhat "trivializing" in a sense that nothing interesting about computers really seems to be continuous. Indeed, if we are looking for a principled or "categorical" and in fact, even a "practical" difference in which to ground a distinction between brains and conventional digital computers, and perhaps even more broadly analog and digital computers, the difference between continuous and discrete physical systems won't get us far.

Despite my broad disagreement I find Maley's contribution very valuable. Due to the many interesting and expressive concepts discussed in (Maley, 2011, 2018a), Maley's framework can serve as a good test piece for comparing intuitions regarding computation and neuroscientific explanations. Such confrontations can often turn out to be very fruitful by allowing us to articulate specific questions which haven't been yet properly addressed in the literature.

It would be interesting to see whether the indeterminacy of type of computers relates to the issue of individuation of computation (Dewhurst, 2016) or how do DAC/ADC figure into the debate on the distinction between analog and digital computers. Perhaps even more importantly, the notions of representational format and resolution should be explicated further and independently of particular definitions of types of representations. Lastly, the role of neuronal noise and its empirically demonstrated benefits await serious philosophical treatment, pointing at a potentially rich interplay between philosophy of computation and neuroscience.

4 From neural analogs to analog computers

There is no clear agreement amongst the different proponents of brains-as-analog-computers view as to what qualifies brains as *analog* computers. Indeed, there is even no clear agreement on what *computers* qualify as analog. This chapter is an attempt at contributing to the debate by considering a "double-inverse" of the question – why do *analogs of brains* qualify as computers?

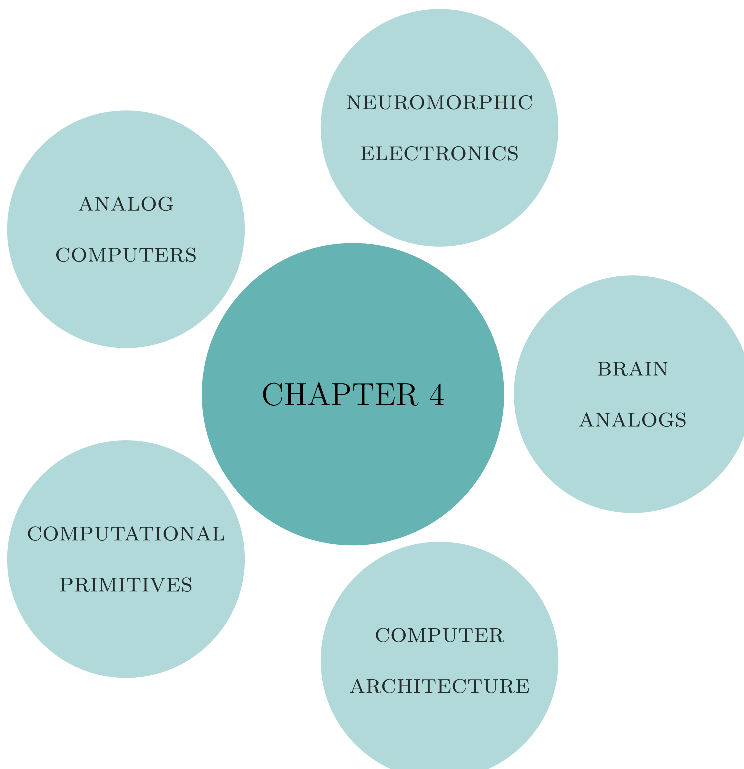

Figure 6: Graphical summary of Chapter 4.

4.1 Introduction

There has been an ongoing debate amongst those agreeing that brains are some kind of computers. The main difference in opinions comes down to what kind of computers does one take brains to be. I'm particularly interested in accounts describing brains as analog computers. However, even these come in a variety of flavors, as there's little agreement on what qualifies a computer as "analog". The debate can be roughly divided into two camps. Curiously enough, both ideas can be seen as von Neumann's legacy.

ANALOG-AS-CONTINUOUS	ANALOG-AS-ANALOGOUS
von Neumann (1948)	von Neumann (1948)
von Neumann (1958)	Lewis (1971)
Eliasmith (2000)	Maley (2011)
Katz (2008)	Shagrir (2010)
...	...

Figure 7: Two accounts of analog computers.

The "analog-as-continuous" view is self-explanatory – according to it a physical system is an analog computer, if it's a computer and it computes with a continuous variable. This leaves us with the "analog-as-analogous". I'll use von Neumann's discussion of "analogy machines" as an entry point. According to von Neumann, an analogy machine is built by following *an* "analogy principle" which is just an instance of the design pattern by which numbers are represented by physical quantities:

> A computing machine may be based on the principle that numbers are represented by certain physical quantities. [...] Operations like

> addition, multiplication, and integration may then be performed by finding various natural processes which act on these quantities in the desired way. (von Neumann, 1951, 293)

While this characterization in no sense implies continuous variables (Lewis, 1971; Maley, 2011), von Neumann nonetheless assumes that analog machines operate in the domain of reals (see (Beebe, 2018) for further discussion). For now, I'd just like to observe that such systems can be either "found" in nature or engineered, but in both cases they're *physical* objects.[38] This is where another notion of analog computers comes into play. Beebe (2018), following Ulmann (2013), talks about analog computation as "computing with models". The basic idea is that computation is somehow done by "modeling" the problem through a construction of a physical analog.[39] While I am interested in analog computers that are "simulating" a certain physical system, I think Beebe (2018) tells only a part of the story. I intend to extend this account to develop a more domain-specific notion of analog computers, as to how it applies to engineered brain analogs.

It is important to note that by using the verb "to simulate" I do not commit to any particular philosophical concept. While I readily agree that both analog simulations and analog computations make use of the same physical devices, I will argue that analog simulation in a narrow sense constitutes only one out of four different types of analog computers. Whereby "type" of an

[38] The question is whether observing such a system out in the wild necessarily leads to a conclusion that it is computing (see (Piccinini, 2017)). I will only discuss engineered systems with an intended use, thus I find such concerns rather toothless, as answering them is trivial – if something is being used as a computer, then it's a computer.

[39] This is in contrast with Shagrir (2010) who defines analog *representations* (rather than the computer itself) as models of environment. Accordingly, an analog computer is a computer manipulating analog representations (cf. Maley (2011)).

analog device is determined by its respective use *and* its user's epistemological attitude. Rather, I use the word as a placeholder for some kind of semantic theory of analog computation.

Moreover, I am not explicitly interested in brains and I will not be addressing the question of whether brains are (analog) computers directly. Rather, I'll talk about particular practice of building *physical analogs* in neuroscience by focusing on a well-defined group of engineered brain-inspired artifacts also known as "neuromorphic" electronics (Mead, 1990; Douglas et al., 1995; Schuman et al., 2017). These devices are built with intention to mimic brains and are often called brain-inspired computers. My goal is thus simple – pursue a less controversial, yet closely related topic by asking how are neuromorphic devices used as analog computers?

4.2 From analogs to analog computers

I'll start by analyzing the seminal programmatic text from Carver Mead, the pioneer of neuromorphic electronics. The purpose of this section is to demonstrate fruitfulness of analyzing neuromorphic electronics following *a* definition of analog computers conforming to the analog-as-analogous view. Particularly, von Neumann's analogy principle seems like a good first fit:

> [W]e should be able to build entire systems based on the organization principles used by the nervous system. I will refer to these systems generically as *neuromorphic systems*. We start by letting the device physics define our elementary operations. These functions provide a rich set of computational primitives, each a direct result of fundamental physical principles. [...] [T]he real trick is to invent the representation that takes advantage of the inherent capabilities of the medium, such as the abilities to generate exponentials,

> to do integration with respect to time, and to implement zero-cost addition using Kirchoff's law. (Mead, 1990, p. 1631)

The main idea is that analogs, and thus analog computers, exhibit similar physical behavior as the physical systems of which they're analogs of. This is in contrast with previous accounts defining analog computers at the level of representation. For example, Shagrir (2010) defines a computational device as analog if it preserves a certain *functional* relationship between representations and represented objects:

> Another way to put it is to say that the representation function is a sort of isomorphism with respect to the functional relation f. Let f be the functional relation between the representing states x and y, namely $f(x) = y$. Let i be the representation function, which maps a representing state to a represented feature. To say that a system computes in the analog sense is to state that functional relations between $i(x)$ and $i(y)$ is also f. (Shagrir, 2010, p. 272)

As later discussed by Shagrir (2010), relationships between x and y, and $i(x)$ and $i(y)$ need not be described by the same mappings. Rather, some kind of formal similarity would suffice. While it is unclear what kind of morphism is implied between f and g, given that $f(x) = y$ and $(g \circ i)(y) = i(y)$, it seems likely that the condition will be trivially satisfied for an arbitrary computational device, when g will correspond to one of device's elementary operations, or when we can represent (some of) the "representing states" on the same device.[40] It think this shows the requirement to be too permissive. Whether this is a legitimate worry or not, Shagrir's proposal doesn't make the cut already

[40] Think of an addition of binary numbers on a digital computer.

due to a much simpler reason. The analogy between neuromorphic systems and biological neural systems lies in their similar physical behavior:

> The significance of neuromorphic systems is that they offer a method of exploring neural computation in a medium whose physical behavior is analogous to that of biological nervous systems and that operates in real time irrespective of size. (Douglas et al., 1995, p. 255)

> The efficiency of neuromorphic analogue VLSI (aVLSI) rests in the power of analogy, the isomorphism between physical processes occurring in different media. (Douglas et al., 1995, p. 258)

A similar worry possibly applies to (Beebe, 2018) as well, insofar as only a functional or representational relationships are considered in his account of "model-based computers".[41]

> A model-based computer is a device which may have a malleable internal structure, and which can represent aspects of the class of problems it is used to solve. The representations should be sufficient to form a model of the target problem class. Under proper use, the organs in the device can be interpreted by the model to function in a manner that we take to solve the target problems. This may or may not be consistent with our understanding of the target problem class. (Beebe, 2018)

What is meant by the "model of the target problem class" or "internal

[41] Considering Beebe's four-way distinction between computers and simulations, which are either analog or model-based, it would seem that only analog simulations require analogous physical behavior.

structure" of a computer? Perhaps, these terms will become clearer if we also look at the notion of an analog computer presented in (Ulmann, 2013), which Beebe (2018) intends to capture as a special case of a "model-based computer":

> An analog computer on the other hand is based on a completely different paradigm: Its internal structure is not fixed — in fact, a problem is solved on such a machine by changing its structure in a suitable way to generate a *model*, a so-called *analog* of the problem. This analog is then used to *analyze* or *simulate* the problem to be solved. Thus the structure of an analog computer that has been set up to tackle a specific problem represents the problem itself while a stored-program digital computer keeps its structure and only its controlling program changes. (Ulmann, 2013, p. 2, as cited in Beebe (2018))

I'll consider two possible understandings of what such model is meant to be, relating to the difference of how the term is used in control theory and (philosophy of) science. A "model" might be just an (approximate) mapping, a mathematical description, of how inputs to the "modelled" system relate to its outputs. In control-theoretic terms, a model corresponds to the transfer function describing the system block in the control diagram. This engineering notion of a model is far weaker than its scientific or philosophical counterpart of a model as a *representation* (of either "a selected part of the world" Frigg and Hartmann (2018) or a scientific theory). In particular, it is weaker insofar as no *epistemic* function is assumed to be performed by a control-theoretic model. That is, a system under control might be treated as a complete black box. There's no need for an understanding or a theory (broadly construed) of *how* or *why* the mapping holds (at least approximately). On the other hand, scientific (or philosophical) understanding of a "model" as an epistemic tool entails a model is a some-kind of *intentional* object used to reason about whatever it is a model of. Importantly, both interpretations entail that there's a specific role to be fulfilled by the model. Namely, to facilitate reasoning about the target system or to solve a computational problem related to it (for example, compute its behavior or a stable state for given initial state variables).[42]

Thus there's a particular reason for why both of these accounts fail at adequately describing the scientific and engineering endeavor behind the design and usage of neuromorphic electronics. Neither Ulmann (2013) nor Beebe (2018) distinguish between the target of analogy and the target of computation.[43] Although I agree that these two often coincide. That is, an analog device is commonly built following an analogy with a certain physical system in order to reason about that very same system, or to solve a (computational) problem related to it. However, it seems that building a "brain-style analog computers" doesn't necessarily serve a purpose of reasoning about brains or solving a related problem. In fact, one of Mead's primary motivations for building neuromorphic computers was energy efficiency:

> For many problems [...] biological solutions are many orders of magnitude more effective than those we have been able to implement using digital methods. (Mead, 1990, p. 1636)

For this reason alone it is worth divorcing the notion of the "target of

[42] Where the latter is presumably less committing in terms of the required richness of representations and the epistemic attitude.

[43] Although to be fair, Ulmann (2017) does distinguish between "direct" and "indirect" analogies: "In short, a direct analogy has its roots basically in the same physical principles as the corresponding problem [...] If the physical principles underlying the problem and analog computer differ, this is called an indirect analog computer." (Ulmann, 2017, p. 2). Nonetheless, Ulmann only focuses on the "indirect analog computers" for the remainder of the book.

computation" from the notion of the "target of analogy".[44] I choose to remain agnostic about whether an analog device must represent the physical system of which it is the analog of, regardless of the application. It does seem to me however, that whether the representation "is there" or not is often irrelevant. Importantly this is not to be confused with the claim that there are no representations involved in using neuromorphic devices.[45] For example, one need not talk about representation of biological retina when considering the functioning and application of its silicon counterpart.

In the next two sections I will first develop a notion of "physical analogy" that could be potentially used to characterize the relationship between the brains and neuromorphic devices. I will then propose a categorization of four different "targets of application" related to the use of neuromorphic devices, with the idea that they might apply more generally to a broader class of analog computers.

4.3 Physical analogies

Let $\mathcal{P}$ be a physical process[46] such that it can be described by a set of differential equations $\mathbf{D}$. That is, $\mathcal{P}$ is treated as a set of tuples p_i that contain values of independent variables such that the equations in $\mathbf{D}$ are simultaneously satisfied. Furthermore, a set of measures $\Theta = \{\theta_i \mid \theta_i : \mathcal{P} \mapsto \mathbb{R}\}$ is defined for the physical quantities whose behavior is described by these equations. The measures are then used to define the syntac-

44 Perhaps it is even more accurate to talk about "target application" instead of "target of computation".

45 It is worth noting that Mead (1990) talks about representations of information, but not about representations of the targets of the analogy.

46 I am using the terms "process" and "system" interchangeably.

tic function g from physical states of $\mathcal{P}$ to a set of computational states, $g : \mathcal{P} \mapsto \mathbf{C}$. This function is a composite of possibly multiple measures and some "pragma" function f, so that $g = f \circ \Theta^n$, $n \geq 1$.[47]

I will start by only assuming a weak constraint for what it means to be an analog. Let $\mathbf{D}_a$ stand for a set of differential equations that describes a physical process that is an analog of a system described by $\mathbf{D}$.[48] Then an isomorphism must exist between $\mathbf{D}$ and $\mathbf{D}_a$.[49] This is intended to capture the necessary condition of "sufficiently" similar physical processes occurring both in the target system and its analog. However, an isomorphism between sets of differential equations, that is, *mathematical descriptions of* physical behavior of a target system and its analog proves to be too weak of a requirement still. Strictly speaking there's no principled distinction between the behavior of ordinary digital or (early) neuromorphic electronics.

> In addition to providing gain, an individual transistor computes a complex nonlinear function of its control and channel voltages. That function is not directly comparable to the function that synapses evaluate using their presynaptic and posynaptic potentials, but a few transistors can be connected strategically to compute remarkably competent synaptic functions.[50] (Mead, 1990, p. 1631)

In fact, all of the early and much of the modern neuromorphic elec-

[47] Function f is just abstracting away from the specifics of how the computation is defined with measures. In other words, f is just a placeholder for a theory, or rather *protocol*, of how certain physical systems are *used* for computing.

[48] As a general notational remark, when discussing pairs of variables introduced in the above paragraphs, the ones with an "a" in the subscript relate to a process that is an analog of a system related to the variables without a subscript.

[49] Cf. Dardashti et al. (2015) on the notion of "nomic isomorphism".

[50] [Note that also for building logic gates one needs multiple transistors. – G.Š.]

tronics are built using the same electrical components as those that make up an ordinary digital computer. Yet, one probably wouldn't want to categorize just *any* electrical device as neuromorphic. The issue was already discussed by Neumann (1958) when trying to justify why neurons should be treated as "all-or-none", that is switching, devices ("organs") with only two relevant states (in comparison to how transistors or vacuum tubes are used in digital computers).

> None of these is an exclusively all-or-none organ (there is little in our technological or physiological experience to indicate that absolute all-or-none organs exist); this, however, is irrelevant. By an all-or-none organ we should rather mean one which fulfills the following two conditions. First, it functions in the all-or-none manner under certain suitable operating conditions. Second, these operating conditions are the ones under which it is normally used; they represent the functionally normal state of affairs within the large organism, of which it forms a part. Thus the important fact is not whether an organ has necessarily and under all conditions the all-or-none character-this is probably never the case-but rather whether in its proper context it functions primarily, and appears to be intended to function primarily, as an all-or-none organ. (Neumann, 1958, p. 296)

Trying to make von Neumann's appeal to "common sense" more specific, the idea is to look only at those dynamics that matter for computation/application. Thus one might want to assume some relationship holds between the measurements related to the two physical systems, Θ and Θ_a in addition to an isomorphism between the descriptions of their behavior. But perhaps we should first ask what kind of measurements are we talking about? For one, "reading" the amount of charge off of on an electrical node is a measurement. As is being sensitive to the time differences between two different events of charge being dropped on a node.[51] One might recognize that the relevant measurements are such that they have an effect on the overall physical behavior of the device in question. That is, the relevant measurements are not external to the functioning of the device and should thus be already described by **D**.

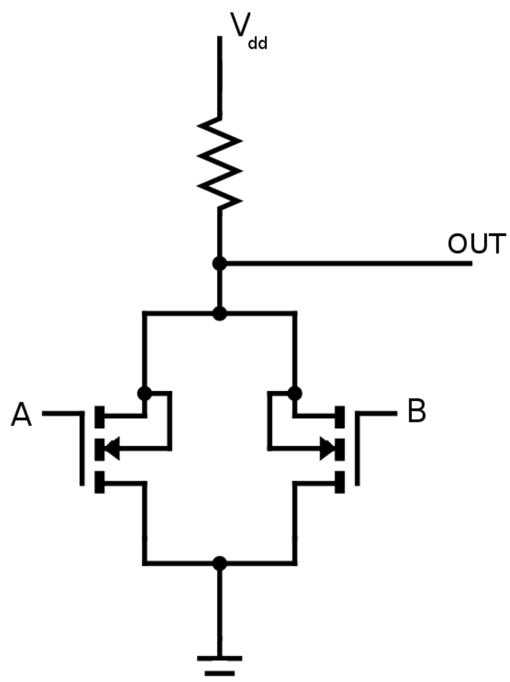

Figure 8: A simple NMOS implementation of a NOR circuit.

For example, in order to constrain a measure to the set {HIGH, LOW},

[51] Importantly, such "measurements" do not imply any intentional states.

pull-down or pull-up resistors[52][53] are commonly used in digital electronics. The argument is that a proper description of a physical system would take into account the effects of these resistors. Thus the isomorphism between $\mathbf{D}$ and $\mathbf{D}_a$ already implies an isomorphism between Θ and Θ_a. Lastly, one could just bite the bullet and agree that transistors used in digital electronics are "potential" neuromorphic devices. The main point being that they could *in principle* be used to build physical analogs of neural systems, but they're not. This leads naturally to the next question – what are (neuromorphic) analogs used for?

4.4 Targets of computation

Neuromorphic computers and analog computers in general are built in order to compute certain functions. This is a commonly (and arguably the only) discussed use case in the philosophical literature (see Beebe, 2018; Shagrir, 2010, and the references therein). While there's not much more to be said about using analogs for computing, I'll take the opportunity to illustrate a more general point. If we want to use something, we must be able to *control* it. For example, in case of a computing device, I want to be able to specify the input, so that

[52] The unnamed resistor in Fig. 8 is a simple example of a pull-up resistor. When neither of the NMOS transistors (representing inputs A and B) is active, the resistor provides a path to output, thus "pulling-up" the output to a high voltage. When one of the transistors is active, the output of the circuit is grounded, that is, in the "low" state. Thus, the output is expected to be only in one of the two states.

[53] Strictly speaking, the pull-down/up resistors or other similar techniques do not restrict the output to a fixed *physical* state. The voltage is still fluctuating and moreover, it does not change instantaneously, but The "HIGH" and "LOW" are computational states – the role of the resistors is to bring the physical states in a certain (stable) dynamical region of the components behavior – within certain time, thus the measures should probably also be described as functions of time (but exhaustively likely as difference rather than differential equations, as discussed in the previous chapter)

the function is computed for a specific value I'm interested in.[54]

However, since the conception of neuromorphic devices computation has been only one of the motivations behind this engineering endeavor. Often, a neuromorphic computer is built to compute something that can be computed with a sufficient and perhaps even better resolution on a conventional (digital) computer and under other reasonable constraints. Moreover, a neuromorphic device can be used to compute something other than what the target system computes or even what it could be used to compute in principle.[55] The purpose of the analogy-guided design is thus not to emulate target's computational capacities but rather other (physical) properties. At least historically, the most common such property is energy efficiency.

> The unavoidable conclusion, which I reached about ten years ago, is that we have something fundamental to learn from the brain about a new and much more effective form of computation. (Mead, 1990, p. 1630)

In all fairness, efficiency concerns have to do with the *efficiency of computation*. Thus, it is not clear how analogs (or specifically, neuromorphic devices) used for "computing" could be distinguished categorically from the ones used for "computing efficiently". The most probable answer is that there is no such clear-cut distinction. Looking at some of the commonly cited benefits of neuromorphic devices over conventional digital computers (Schuman et al., 2017), one could argue that there's a difference between "performing a computation more efficiently" (real-time performance, parallelism, speed, fault

[54] In this sense, one might argue that the fact that an analog of a certain system is a computer does not imply that the target system is a computer itself. The common reason being simply that it cannot be *used* as a computer, insofar the control over it is lacking.

[55] This is what Beebe (2018) and Ulmann (2013) miss in their analysis.

tolerance) and "performing *cheaper* computation" (scalability, low power, less silicon area). However, these two are hardly ever considered separately, but are usually cast in terms of trade-offs.[56] In fact, we might be even talking about *the same device*. That is, the distinction is likely to be contextual and based on the difference in the epistemic attitudes of their users.

Building devices that compute the same or a similar function as a target neural circuit, by emulating its physical behavior has proven instrumental both in identifying general computing strategies *and* the mechanisms performing the computation. Mechanism discovery is arguably the most important aspect of using neuromorphic devices for explanations in neuroscience.[57]

> The structure executing this level-normalization operation performs many other functions as well, such as computing the contrast ratio and enhancing edges in the image. Thus, the mechanisms responsible for keeping the system operating over an enormous range of image intensity have important consequences with regard to the representation of data. (Mead, 1990, p. 1632)

> By designing neuromorphic systems, we enlarge our vocabulary of computational primitives that provide a basis for understanding computation in nervous systems. (Douglas et al., 1995, p. 279)

Once a plausible mechanism is identified, further inference can lead to more functional descriptions of what computation is being performed. The inferred abstractions or algorithms are valuable because they might suggest a

[56] For example, in order to increase the speed of the computation more silicon area and probably also more power will be needed.

[57] In fact, some authors (Craver, 2009; Kaplan, 2011) have argued that *all* explanations in neuroscience are *mechanistic*, although see (Chirimuuta, 2018, 2014).

way to adapt the solving strategy to either solve similar, but distinct problems or to design a different mechanism implementing the same functionality (Ullman, 2019).

> The center-surround computation sometimes is referred to as a Laplacian filter, which has been used widely in computer vision systems. This computation, which can be approximated by a difference in Gaussians, has been used to help computers localize objects; this kind of enhancement is effective because discontinuities in intensity frequently correspond to object edges. Both of these mathematical forms express, in an analytically tractable way, the computation that occurs as a natural result of an efficient physical implementation of local normalization of the signal level. (Mead, 1990, p. 1632)

> The biological relevance of the chip is that it expresses the stereo-fusion problem as just one instance of a general class of constrain-satisfaction problems in sensory perception and shows how this class of problems can be [sic] computed with neuron-like elements. (Douglas et al., 1995, p. 277)

Non-neuromorphic physical analogs have been largely discussed in regard to one other type of scientific practice. Namely, use of analog simulation in physics (Dardashti et al., 2015, 2019; Thébault, 2019). While Beebe (2018) considers analog simulation as a special case of analog computation, the wording in Ulmann (2013) could imply the opposite, namely that analog computers are a special case of analog simulations. For the purposes of this here discussion, it is sufficient to think of (analog) simulation as an epistemological practice that performs a function similar to that of an experiment – that is, gathering *data* to either verify or formulate a hypothesis.

> This chip highlights an interesting methodological point–because neuromorphic models are constrained by what can be implemented in a physical medium, they can provide insight into biological design.(Douglas et al., 1995, p. 264)[58]

Admittedly, when it comes to neuromorphic engineering, the difference between analog simulation and mechanism (or algorithm) discovery becomes blurred. While the latter are intended to inform the design of devices that have application beyond "knowledge production", the former is most often geared towards confirmation of a scientific theory (Utagawa et al., 2011; Douglas et al., 1995) or verification of methodology (Jonas and Kording, 2017). It seems reasonable to expect that the same research activity will be often in service of both pursuits. The more important distinction is that between "analog simulation" or "analogue emulation"[59] on one side, and digital simulation on the other.

The difference between digital simulation and emulation[60] is that a

[58] The quote continues "The response of the basilar membrane scales–that is, the spatial pattern of the response is invariant with frequency except for a displacement along the membrane. There are two physical models that give rise to scaling. In the first–constant mass scaling–the mass of the membrane and the density of the fluid are constant, but their stiffness changes exponentially along its length. In the second model–increasing mass scaling–all three change exponentially along the length of the membrane. Although the behavior of these two models is indistinguishable, their implications for physical implementation are radically different. The constant mass model requires that membrane stiffness change by a factor of about one million. Such a range of variation can be simulated on a digital computer, but it cannot be implemented easily in a physical device. This suggests that the increasing mass model, in which the range of variation can be absorbed by three parameters rather than one, should be adopted."

[59] I hold it that the notion of "analog simulation" as used in herein and by Dardashti et al. (2015), Beebe (2018), and Ulmann (2013) is synonymous with the notion of "analogue emulation" that is more often used in the NE literature.

[60] For the sake of simplicity, I will refer to digital simulation only as "simulation" and to

simulation does not necessarily conform to the same environmental constraints as the emulated system. To make this intuition a bit more precise, think for example, of a physical system $\mathcal{P}$ that is subject to a certain physical law $\mathbb{L}$, such that its behavior is described by a set of equations $\mathbf{D}$. Let $\mathcal{S}$ and $\mathcal{E}$ be a simulation of the $\mathcal{P}$ and a system emulating it, respectively.

Assume now that a transition from two states, p_i to p_k can only occur by the system $\mathcal{P}$ assuming a set of intermediate states. For any such member of this set p_j, such that it can be both simulated and emulated by the $\mathcal{S}$ and $\mathcal{E}$, let s_j and e_j be the states that simulate and emulate p_j, respectively. A simulation might be set up such that the transition occurs directly from the representation of the states p_i to p_k, that is from s_i to s_k, however, an emulation of the system will always follow the same transition of the states $p_i, ..., p_j, ..., p_k$, namely by the system $\mathcal{E}$ assuming the sequence of states $e_i, ..., e_j, ..., e_k$. While the behavior of $\mathcal{P}$, $\mathcal{E}$, and $\mathcal{S}$ can be all *described* by $\mathbf{D}$, invocation of the physical law $\mathbb{L}$ is sufficient to properly *explain* the behavior of $\mathcal{E}$, but no such constraint holds for $\mathcal{S}$.

> The specialized but efficient nature of neuromorphic systems causes analogue emulation to play a different role in the investigation of biological systems than does digital simulation. Analogue emulation is particularly useful for relating the physical properties of the system to its computational function because both levels of abstraction are combined in the same system. In many cases, these neuromorphic analogues make direct use of device physics to emulate the computational processes of neurons so that the base level of the analysis is inherent in the machine itself. Because the computation is cast as a physical process, it is relatively easy to move from emulation to

analog emulation/simulation as "emulation" in the present paragraph.

physiological prediction. (Douglas et al., 1995, p. 278-9)

Consider the below summary of the 5 different use cases of physical analogs discussed in this section. If a physical system $\mathcal{A}$ is built following an analogy principle, so that it acts as a physical analog of another system $\mathcal{P}$, and $\mathcal{P}$ can be described as computing a function f, then I might use $\mathcal{A}$ to:

(A0) *control* the $\mathcal{A}$ as to (approximately) measure f for a specified "input" $f(i)$, or

(A1) learn about a general class of "computing strategies" for a given class of problems, or

(A2) identify particular mechanisms of $\mathcal{P}$ responsible for "implementing" f, or

(A3) replace $\mathcal{B}$ with $\mathcal{A}$ to *reimplement* f more "efficiently"[61], or

(A4) gather *data* about $\mathcal{A}$ that can be used to reason about a hypothesis regarding $\mathcal{P}$.

The idea at its simplest is that only by properly emulating $\mathcal{P}$, rather than merely simulating it, can $\mathcal{A}$ be used for identification of mechanisms (A2) and hypothesis formulation or testing (A4). In order to make this distinction more tangible it is worth comparing neuromorphic computers to conventional digital computers, given that the two are presumably used to emulate and simulate the brain, respectively. Thus, the next section considers a number of ways neuromorphic computers diverge from their digital counterparts in order to remain more faithful to the structure and dynamics of neural circuits.

[61] Where $\mathcal{B}$ is just another physical system that can be used to compute f and efficiency doesn't need to relate to computation.

4.5 Computing with noise & time

A great deal of the discussion in this section could certainty be cast in more broader terms of analog and digital computers and differences between the two. Indeed, the working hypothesis since the beginning of the chapter states that neuromorphic electronics fall under the category of analog-as-analogous computers. Nonetheless, the first principal distinction is domain specific insofar the brain inspired design of computers has been regularly described as an alternative to a von Neumann architecture (Walter et al., 2015; Schuman et al., 2017; Boybat et al., 2017; Gkoupidenis et al., 2017).

> Neuromorphic computing has emerged in recent years as a complementary architecture to von Neumann systems. (Schuman et al., 2017, p. 1)

Clearly, "to be an alternative" is not meant in the same sense as, for example, when discussing a choice between von Neumann and Harvard architecture. Indeed, in this case both of the latter would be lumped together under the label "von Neumann architecture" which became a common designate for sequential and modular design.

> It is now well recognized that electronic circuits based on traditional von Neumann architerctures are not well-adapted to capture the real world information processing capability of biological nervous systems. The main reason of this limitation is the so-called von Neumann bottleneck, due to the physical separation of computational and memory units. (Gkoupidenis et al., 2017, p. 2)

> [I]t seems paradox that simulating a tiny fraction of the human brain is only possible at a multiple of its power consumption. This can

> be at least partly explained by the completely different paradigms underlying standard Von Neumann CPUs on the one hand and neural networks on the other. While the former implement a sequential model of computation which is based on a centralized local storage, information processing in the latter is massively parallel and distributed. (Walter et al., 2015, p. 153)

The bottleneck could in principle be overcome by simply moving the memory to the same chip as the CPU[62], presumably also in combination with using new memory technologies and under revised standards (for example, specifying lower retention times) (Wong and Salahuddin, 2015). However, one might also want to rethink the strong separation between processing and memory altogether. As a rule of thumb, a conventional computer uses memory for two main types of *data* – the input and output of the computer, and the set of instructions that determine the exact program that is being executed.[63]

In principle, there are ways of working around both these needs. Interestingly enough, they have to do in part with the assumption of sequential processing. If all the data is processed in parallel, there's no more need to hold the-yet-unprocessed data chunks in memory waiting for its turn (Ulmann, 2017). Moreover, by embedding the program *within the computer structure*, rather than delegating control to the memory[64] stored instructions, the need for program memory is circumvented as well.[65] By mimicking biological neural

[62] Or *vice versa*, by adding some low-overhead processing capacity to the memory, as suggested by the processing-in-memory (PIM) approach (Zhang et al., 2013).

[63] The latter basically being a physical instance of the stored-program concept invented by Turing in his seminal paper (Turing, 1936). See Pelaez (1999) for a more in-depth historical perspective.

[64] Cf. the discussion by Neumann (1958) on the difference between "memory-stored" and "plugin" control.

networks, neuromorphic electronics implement "distributed memory" or somehow otherwise *emergent* mnemonic capacity by making use of strong recurrent connectivity (Hunt and Hayden, 2017), electro-chemical global coupling (Gkoupidenis et al., 2017), or a winner-takes-all protocol for physically emulated synaptic competition (Manning et al., 2018).

Time- or rather timing-dependent computation is a good illustration of this point. Conventional digital computers commonly implement all sorts of clocking and timing functionalities. However, a sensitivity of a program to, for example, a difference in the timing between occurrences of two separate events is a matter of logic, presumably some branching instructions. A conventional digital computer will stay the same after executing the program, whereas computationally relevant timing differences will have an effect on neurons (*e.g.*, due to a LTP or LTD, see Footnote 27). That is to say that a *physical* and *computational* state of such a computer can be divorced[66], whereas the same couldn't be said for a biological or emulated neural network. Referring back to Section 4.3 it seems sensible to speculate that physical analogs differ from conventional digital or more precisely, stored-program computers, in that the set of measurements Θ used to describe the latter is invariant, whereas with the former it is subject to change.

As the name of the section would suggest another difference between the program-stored and analog-as-analogous computers has something to do

[65] Importantly though, the value of program-stored computers lies exactly in this structure invariance. Much like a universal Turing machine is able to emulate any other Turing machine, stored-program computers are *general* purpose. While this is not to claim that analog-as-analogous computers serve a single-purpose computation, it would certainly seem that they're less adapt at performing different computation without being restructured at an expense of likely significantly greater amount of resources.

[66] Indeed, this seems to be a necessary condition for having a general-purpose or reprogrammable computer.

with noise. While not as strongly represented in the literature as the analog-as-continuous and analog-as-analogous accounts, there's also a third characterization of analog computers[67], namely, as "approximate" – noisy or imprecise – procedures (Haugeland, 1981; Katz, 2008). Perhaps not too surprisingly, the original thoughts on the matter can also be attributed to von Neumann (1951).

> [T]he critical question with every analogy procedure is this: How large are the uncontrollable fluctuations of the mechanism that constitute the "noise," compared to the significant "signals" that express the numbers on which the machine operates? The usefulness of any analogy principle depends on how low it can keep the relative size of the uncontrollable fluctuations-the "noise level." (von Neumann, 1951, p. 293)

However, in accordance to the discussion in Section 4.3 about there hardly being a qualitative difference between the physical descriptions of digital or analog computers, one might wonder if the same considerations would apply to the presence of noise. As also discussed by von Neumann (1951), there's also error in digital procedures – for example, it is commonplace to talk about "machine epsilon", that is, the upper bound on the rounding error in floating point arithmetic.[68]

The important difference between the noise level of a digital machine,

[67] Strictly speaking, Haugeland (1981) and Katz (2008) talk about the difference between analog and digital representations or "procedures". In light of the discussion in Chapter 3, the leap to a classification of computers or devices is hardly controversial; even more so when considering that the references are here for the purpose of the narrative rather than an argument.

[68] For example, an error is almost inevitable when performing a division of two different numbers and with the denominator not being a power of the base in which the numbers are represented.

> as described above, and of an analogy machine is not qualitative at all; it is quantitative. (von Neumann, 1951, p. 295)

Another commonality between the two engineering approaches would seem to be that in both cases noise is perceived as a quantity that needs to be minimized. As per the classification schema in Section 1.2 it is apparent that, at least at the years of its inception and early development, the field of neuromorphic engineering remained entrenched in the "engineering paradigm", by following a positive (applicative) analogy with the conventional electrical circuits.

> Biological systems appear to make good use of noise in diverse processes [...] This contrasts with engineering where noise is usually considered as a disturbance. (Knuuttila and Loettgers, 2013, p. 164)

> Noise is a form of redundancy because it does not supply additional information about the world, so bandwidth is wasted by transmitting it. (Douglas et al., 1995, p. 262)

What makes neuromorphic engineering a particularly interesting case study for philosophers and historians of science is that with the rising recognition of the importance of noise in neural circuits (see 1.2), the engineering filed has undergone a similar paradigm shift. For example, novel neuromorphic circuits have exploited noise for pulse density modulation (Utagawa et al., 2007, 2011), synchronization between isolated circuits (Utagawa et al., 2008), stochastic resonance (Gonzalez-Carabarin et al., 2014), and coincidence detection (Oya et al., 2006).

Finally, the term "architecture" is usually an abbreviation of "instruction set architecture". However, it would hardly seem sensible to talk about "instructions" when it comes to neuromorphic computers. Related to that, talking about the difference in *design* might be a bit of a misnomer insofar as state-of-the-art neuromorphic electronics are concerned, just like brains and other biological systems are not "designed", but rather evolved (Zhu et al., 2019). Considering the noise in terms of random fluctuations and perturbations that are constitutive to the development of biological systems (Chirimuuta, 2017), we might speculate that a lot of such "beneficial noise" is in fact mischaracterized endogenous activity (Bechtel, 2012). Conceptually this should serve as another example against characterizing analogical relationships (*e.g.*, between brains and neuromorphic computers) exclusively in terms of I/O, or functions broadly conceived.

4.6 Conclusion

In this chapter I focused on a rather under-explored question in philosophy of computation. The idea is to provide a rational reconstruction of scientific *and engineering* practices falling under the broader category of "neuromorphic computing". I see this as a valuable contribution to the ongoing debate about computationalism about brains and conception of "analog computers". Importantly, the value of the contribution is not derived antagonistically by replacing other existing theories. In fact, I tried to be as agnostic as possible, to rather provide a kind of template that has yet to be filled out to fully specify a philosophical account. As far as the applicability of the discussed theory of analog computers to questions concerning computationalism about brains, I find it suitable to quote from Beebe (2018):

> Under this framework, we would answer 'yes' to the question of whether the brain is a model-based computer, and also 'yes' to the

> question of whether brain processes are computational. However, this may be a bit premature since we have noticed that computation is dependent on a user–and what would be using this model-based computer? (Beebe, 2018)

The only difference being that I'd be more willing to conclude that brains are *not* analog computers in the sense outlined here. This unveils a broader characteristic of the framework. If system A is a physical analog of a system P, and system A is used as an (analog) computer in a certain sense, it doesn't not follow that P is also an analog computer. It would be interesting to see whether an argument could be made as to whether this kind of symmetry or asymmetry should be a required desiderata for a definition of analog computation.

In complement to the various shortcomings of the distinction between continuous and discrete physical systems discussed in the previous chapter, the analysis of neuromorphic computers offers a positive reason for grounding the definition of analog computers in a theory of (physical) analogies. Secondly, it would appear that the current analog-as-analogous literature is entrenched in thinking in terms of I/O properties. The literature on neuromorphic electronics offers a handful of examples that suggest a more broader understanding of what makes up an analog computer. Importantly, many of these examples come from a clever use of brain analogs' non-deterministic behavior akin to neural noise.

5 Conclusion

Random fluctuations in neural activity, commonly described as neural noise have been repeatedly dismissed as meaningless and even deemed detrimental to the performance of the neural system. Recent advances in neuroscientific research suggest the opposite. There are good reasons to believe that at least a part of the noise plays a central role in the functioning of the neurobiological systems. On the factual level this goes beyond "classical" stochastic resonance and might in fact often have to do with processes that are not easily described in reference to signals and information, or even more generally in terms of inputs and outputs. It was speculated that this methodological incompatibility is the reason for the historical treatment of noise as a redundant physical quantity and assuming that evolutionary forces push towards its minimization.

This thesis considered the flip side of the coin, emphasizing conceptual rather than factual roles played by neural noise. Starting with the first manuscript we need to take a step back. Neural activity described as noise is sometimes just that – noise, a redundant quantity detrimental to the performance of the system in question. As such it is ubiquitous in neural circuits underlying numerous behavioral domains like motor control, perception, and value or preference based choices. Importantly, understanding that this noise can be conceptualized as a source of uncertainty led researchers to describe the three aforementioned behavioral domains using a decision-theoretic framework. Consequently, a number of mathematical models and even experimental paradigms have been shared between the research domains, and established normative claims based on a more faithful comparison and methodological alignment.

Related to this "factual" presence of noise, it was argued before that brains inevitably compute with *discrete* values. This argument is reiterated in the second manuscript. If a variable is noisy and this noise is accounted for, then the system will not discriminate between a number of precise values (in fact *uncountably* many), effectively using a discretized variable. Moreover, special attention was given to the fact that presence of noise is a recurrent *explanandum* of neural miscomputation. Tying the two together amounts to a novel practical argument for why brains shouldn't be conceptualized as continuous computers, if they were to be conceptualized as computers in the first place. By *modus tollens*, asserting that brains compute with continuous variables amounts to denying that brains compute with noisy variables. Therefore, asserting that brains compute with *continuous* variables fails to accommodate a common neuroscientific practice. In as much a conceptualization of neural computation should aim at explicating existing scientific practices and current best theories, the failure to do so serves as a strong reason for rejecting the said conceptual framework.

The third manuscript rounds out the discussion that was started in the general introduction. Neuromorphic devices are presented as an example of physical systems that are deemed (analog) computers on the basis of their *non* I/O properties. Furthermore, I argued that these devices also make up for an interesting case study in the philosophical discussion of noise and its role in (biological) sciences as a "non-paradigmatic" engineering practice, given that the noise has been increasingly more often considered as a design feature, rather than a nuisance. The upshot being that design is, perhaps inevitably, only concerned with I/O relations, but there is value in recognizing some of the noise as an invaluable endogenous activity.

Acknowledgments

I feel indebted to a lot of people. If it wouldn't have been for Borut Cerkovnik and Gregory Wheeler I wouldn't have taken this path in the first place. I am as indebted to Stephan Hartmann as well, I couldn't have wished for a more knowledgeable and caring mentor. Ophelia Deroy and Stefan Glasauer always had time for me and our discussions. Here I'd also like to mention Stephan Sellmaier, he's a cool guy and I've learned a lot from him. I thank you all for your support.

It was a privilege to study at GSN. They supported me financially and logistically throughout my studies. I'd like to thank to all the administration staff at GSN and MCMP, especially Lena, Maria, Nadine, Renate, and Stephanie – you've helped me countless times.

I met many wonderful people and I hope to remember you all. Perhaps writing down your names will help with recalling the nice moments: Adam, Adriana, Anna, Anna Lena, Anki, Aurore, Berni, Cameron, Conrad, Dilectiss, Dylan, Felix, Holly, Ines, Javi, Joe, Johannes, Jure, Luka, Marga, Marleen, Mauri, Mičo, Mónica, Nika, Nora, Ot, Pari, Sachith, Sebastian, Taimur, Tilen, Timo. Thank you! Cameron deserves a special mention. Our countless lengthy discussions have strongly shaped the way I think about topics and issues presented herein, but also many others.

A lot has happened during this time and I have to thank to my smart and beautiful co-op player Mónica for making this game worth playing. Y para la familia en España: gracias por todo!

Družina, hvala za vso podporo. Jaz sem še vedno jaz, papir je pa papir. Na še mnoga leta in trenutke v sončku, ob smehu ali pesmi, pa morda s špricarčkom v roki! Zelo vas imam rad in sem srečen, da vas imam.

Bibliography

Acerbi, L., Vijayakumar, S., and Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS Computational Biology*, 10(6):e1003661.

Bale, M., Campagner, D., Erskine, A., and Petersen, R. (2015). Microsecond-scale timing precision in rodent trigeminal primary afferents. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 35:5935.

Bear, Mark F. & Barry W. Connors, M. A. P. (2016). *Neuroscience: exploring the brain*. Wolters Kluwer, 4th edition.

Bechtel, W. (2012). Understanding endogenously active mechanisms: A scientific and philosophical challenge. *European Journal for Philosophy of Science*, 2.

Bechtel, W. and Abrahamsen, A. (2010). Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A*, 41(3):321 – 333. Computation and cognitive science.

Beebe, C. (2018). Model-based computation. *Natural Computing*, 17(2):271–281.

Bettman, J. R., Johnson, E. J., and Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, 45(1):111 – 139.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and D Cohen, J. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113:700–65.

Bonawitz, E. and Griffiths L., T. (2010). Deconfounding hypothesis generation and evaluation in bayesian models. *Proceedings of the Thirty-second Cognitive Science Society. Austin, TX: Cognitive Science Society*, pages 2260–2265.

Boon, M. (2017). An engineering paradigm in the biomedical sciences: Knowledge as epistemic tool. *Progress in Biophysics and Molecular Biology*, 129:25 – 39. Validation of Computer Modelling.

Boybat, I., Gallo, M., S.R., N., Moraitis, T., Parnell, T., Tuma, T., Rajendran, B., Leblebici, Y., Sebastian, A., and Eleftheriou, E. (2017). Neuromorphic computing with multi-memristive synapses. *Nature Communications*, 9.

Brighton, H. (2019). Beyond quantified ignorance: rebuilding rationality without the bias bias. *Economics Discussion Papers*, (2019-25).

Butts, D., Weng, C., Jin, J., Yeh, C.-I., Lesica, N., Alonso, J.-M., and Stanley, G. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature*, 449:92–5.

Buzsáki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929.

Carandini, M. and Heeger, D. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese*, 191(2):127–153.

Chirimuuta, M. (2017). Crash testing an engineering framework in neuroscience: Does the idea of robustness break down? *Philosophy of Science*, 84(5):1140–1151.

Chirimuuta, M. (2018). Explanation in computational neuroscience: Causal and non-causal. *British Journal for the Philosophy of Science*, 69(3):849–880.

Chrobak, J. J. and Buzsáki, G. (1998). Gamma oscillations in the entorhinal cortex of the freely behaving rat. *Journal of Neuroscience*, 18(1):388–398.

Chua, L. and Kang, S.-M. (1976). Memristive devices and systems. *Proceedings of the IEEE*, 64:209–223.

Craver, C. F. (2009). *Explaining the Brain*. Oxford University Press.

Csicsvari, J., Jamieson, B., D Wise, K., and Buzsáki, G. (2003). Mechanisms of gamma oscillations in the hippocampus of the behaving rat. *Neuron*, 37:311–22.

Dardashti, R., Hartmann, S., Thébault, K., and Winsberg, E. (2019). Hawking radiation and analogue experiments: A bayesian analysis. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 67:1 – 11.

Dardashti, R., Thébault, K., and Winsberg, E. (2015). Confirmation via analogue simulation: What dumb holes can tell us about gravity. *The British Journal for the Philosophy of Science*, 68.

Dekker, T. M. and Nardini, M. (2016). Risky visuomotor choices during rapid reaching in childhood. *Developmental Science*, 19(3):427–439.

Dewhurst, J. (2016). Individuation without Representation. *The British Journal for the Philosophy of Science*, 69(1):103–116.

Douglas, R., Mahowald, M., and Mead, C. (1995). Neuromorphic analogue vlsi. *Annual review of neuroscience*, 18:255–81.

Dutilh, G. and Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review*, 23(3):723–737.

Egan, F. (2010). Computational models: A modest role for content. *Studies in History and Philosophy of Science Part A*, 41(3):253–259.

Eguchi, A., Isbister, J., Ahmad, N., and Stringer, S. (2018). The emergence of polychronization and feature binding in a spiking neural network model of the primate ventral visual system. *Psychological Review*, 125.

Eliasmith, C. (2000). Is the brain analog or digital? *Cognitive Science Quarterly*, 1(2):147–170.

Eliasmith, C. (2010). How we ought to describe computation in the brain. *Studies in History and Philosophy of Science Part A*, 41(3):313–320. Computation and cognitive science.

Ermentrout, B., Galán, R., and Urban, N. (2008). Reliability, synchrony and noise. *Trends in neurosciences*, 31:428–34.

Faisal, A., Selen, L., and Wolpert, D. (2008). Noise in the nervous system. *Nature reviews. Neuroscience*, 9:292–303.

Fechner, H. B., Schooler, L. J., and Pachur, T. (2018). Cognitive costs of decision-making strategies: A resource demand decomposition analysis with a cognitive architecture. *Cognition*, 170:102–122.

Feurra, M., Paulus, W., Walsh, V., and Kanai, R. (2011). Frequency specific modulation of human somatosensory cortex. *Frontiers in Psychology*, 2:13.

Fiete, I. R., Hahnloser, R. H., Fee, M. S., and Seung, H. S. (2004). Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *Journal of Neurophysiology*, 92(4):2274–2282.

Frigg, R. and Hartmann, S. (2018). Models in science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2018 edition.

Fu, W. (2016). The central role of heuristic search in cognitive computation systems. *Minds and Machines*, 26(1-2):103–123.

Genewein, T. and Braun, D. A. (2016). Bio-inspired feedback-circuit implementation of discrete, free energy optimizing, winner-take-all computations. *Biological Cybernetics*, 110(2):135–150.

Gigerenzer, G. and Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1):107–143.

Gigerenzer, G., Dieckmann, A., and Gaissmaier, W. (2012). Efficient cognition through limited search. In Peter M. Todd, Gerd Gigerenzer, T. A. R. G., editor, *Ecological rationality: Intelligence in the world*, pages 241–273. New York: Oxford University Press.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62(1):451–482.

Gigerenzer, G. and Sturm, T. (2011). How (far) can rationality be naturalized? *Synthese*, 187:1–26.

Gkoupidenis, P., Koutsouras, D., and Malliaras, G. (2017). Neuromorphic device architectures with global connectivity through electrolyte gating. *Nature Communications*, 8.

Glöckner, A. and Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica*, 139(3):532–542.

Glöckner, A., Hilbig, B. E., and Jekel, M. (2014). What is adaptive about adaptive decision making? a parallel constraint satisfaction account. *Cognition*, 133(3):641–666.

Gold, J. I. and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, 5(1):10–16.

Gonzalez-Carabarin, L., Asai, T., and Motomura, M. (2014). Application of nonlinear systems for designing low-power logic gates based on stochastic resonance. *Nonlinear Theory and Its Applications, IEICE*, 5(4):445–455.

Hagura, N., Haggard, P., and Diedrichsen, J. (2017). Perceptual decisions are biased by the cost to act. *eLife*, 6:e18422.

Hahnloser, R., A Kozhevnikov, A., and Fee, M. (2002). An ultra-sparse code underlies the generation of neural sequence in a songbird. *Nature*, 419:65–70.

Hämmerer, D., Bonaiuto, J., Klein-Flügge, M., Bikson, M., and Bestmann, S. (2016). Selective alteration of human value decisions with medial frontal tDCS is predicted by changes in attractor dynamics. *Scientific Reports*, 6(1):25160.

Hanks, T. D. and Summerfield, C. (2017). Perceptual decision making in rodents, monkeys, and humans. *Neuron*, 93(1):15–31.

Harris, D. K., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424:552–6.

Haugeland, J. (1981). Analog and analog. *Philosophical Topics*, 12(1):213–225.

Hertwig, R. and Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. *Perspectives on Psychological Science*, 11(3):359–372.

Hertwig, R. and Pachur, T. (2015). Heuristics, history of. In *International Encyclopedia of the Social & Behavioral Sciences*, pages 829–835. Elsevier.

Hunt, L. T. and Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience*, 18(3):172–182.

Hutcheon, B. and Yarom, Y. (2000). Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends in Neurosciences*, 23:216–222.

Izhikevich, E. (2006). Polychronization: Computation with spikes. *Neural computation*, 18:245–82.

Jarvstad, A., Hahn, U., Rushton, S. K., and Warren, P. A. (2013). Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *Proceedings of the National Academy of Sciences*, 110(40):16271–16276.

Jocham, G., Hunt, L. T., Near, J., and Behrens, T. E. J. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nature Neuroscience*, 15(7):960–961.

Jonas, E. and Kording, K. (2017). Could a neuroscientist understand a microprocessor? *PLOS Computational Biology*, 13:e1005268.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93:1449–1475.

Kaplan, D. (2011). Explanation and description in computational neuroscience. *Synthese*, 183:339–373.

Kaplan, D. M. and Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78(4):601–627.

Katsikopoulos, K., J Schooler, L., and Hertwig, R. (2010). The robust beauty of ordinary information. *Psychological review*, 117:1259–66.

Katz, M. (2008). Analog and digital representation. *Minds and Machines*, 18(3):403–408.

Khader, P. H., Pachur, T., Weber, L. A. E., and Jost, K. (2016). Neural signatures of controlled and automatic retrieval processes in memory-based decision-making. *Journal of Cognitive Neuroscience*, 28(1):69–83.

Knuuttila, T. and Loettgers, A. (2013). Basic science through engineering? synthetic modeling and the idea of biology-inspired engineering. *Studies in*

History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44(2):158 – 169. Philosophical Perspectives on Synthetic Biology.

Knuuttila, T. and Loettgers, A. (2014). Varieties of noise: Analogical reasoning in synthetic biology. *Studies in History and Philosophy of Science Part A*, 48:76 – 88.

Lewis, D. K. (1971). Analog and digital. *Noûs*, 5(3):321–327.

Louie, K., Glimcher, P. W., and Webb, R. (2015). Adaptive neural coding: from biological to behavioral decision-making. *Current Opinion in Behavioral Sciences*, 5:91 – 99.

Louie, K., Khaw, M. W., and Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, 110(15):6139–6144.

Louie, K., LoFaro, T., Webb, R., and Glimcher, P. W. (2014). Dynamic divisive normalization predicts time-varying value coding in decision-related circuits. *Journal of Neuroscience*, 34(48):16046–16057.

Magee, J. and Johnston, D. (1997). A synaptically controlled, associative signal for hebbian plasticity in hippocampal neurons. *Science (New York, N.Y.)*, 275:209–13.

Maley, C. (2011). Analog and digital, continuous and discrete. *Philosophical Studies*, 155(1):117–131.

Maley, C. (2018a). Toward analog neural computation. *Minds and Machines*, 28(1):77–91.

Maley, C. J. (2018b). Continuous neural spikes and information theory. *Review of Philosophy and Psychology*, pages 1–21.

Manning, H., Niosi, F., Gomes da Rocha, C., Bellew, A., O'Callaghan, C., Biswas, S., Flowers, P., Wiley, B., Holmes, J., Ferreira, M., and Boland, J. (2018). Emergence of winner-takes-all connectivity paths in random nanowire networks. *Nat Commun*, 9(3219).

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

McDonnell, M. and Ward, L. (2011). The benefits of noise in neural systems: Bridging theory and experiment. *Nature Reviews Neuroscience*, 12:415–425.

Mead, C. (1990). Mead c (1990, october) neuromorphic electronic systems. proc ieee. *Proceedings of the IEEE*, 78:1629 – 1636.

Minev, Z., Mundhada, S., Shankar, S., Reinhold, P., Gutierrez-Jauregui, R., Schoelkopf, R., Mirrahimi, M., Carmichael, H., and Devoret, M. (2019). To catch and reverse a quantum jump mid-flight. *Nature*, 570.

Mousavi, S. and Gigerenzer, G. (2017). Heuristics are tools for uncertainty. *Homo Oeconomicus*, 34(4):361–379.

Neumann, J. v. (1958). *The Computer and the Brain.* Yale University Press, New Haven, CT, USA.

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3):113–126.

Neyedli, H. F. and Welsh, T. N. (2014). People are better at maximizing expected gain in a manual aiming task with rapidly changing probabilities than with rapidly changing payoffs. *Journal of Neurophysiology*, 111(5):1016–1026.

Nowick, S. M. and Singh, M. (2015). Asynchronous design–part 1: Overview and recent advances. *IEEE Design Test*, 32(3):5–18.

Oya, T., Asai, T., Kagaya, R., Hirose, T., and Amemiya, Y. (2006). Neuronal synchrony detection on single-electron neural networks. *Chaos, Solitons & Fractals*, 27(4):887 – 894.

O'Leary, T. and Wyllie, D. J. A. (2011). Neuronal homeostasis: time for a change? *The Journal of Physiology*, 589(20):4811–4826.

Panzeri, S., Brunel, N., Logothetis, N., and Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. *Trends in neurosciences*, 33:111–20.

Parberry, I. (1994). *Circuit complexity and neural networks*. MIT Press.

Pelaez, E. (1999). The stored-program computer: Two conceptions. *Social Studies of Science*, 29(3):359–389.

Philiastides, M. G., Biele, G., and Heekeren, H. R. (2010). A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences*, 107(20):9430–9435.

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. Oxford University Press.

Piccinini, G. (2017). Computation in physical systems. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition.

Piccinini, G. and Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science*, 37(3):453–488.

Piccinini, G. and Scarantino, A. (2010). Computation vs. information processing: why their difference matters to cognitive science. *Studies in History and Philosophy of Science Part A*, 41(3):237 – 246.

Polanía, R., Krajbich, I., Grueschow, M., and Ruff, C. (2014a). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, 82(3):709–720.

Polanía, R., Krajbich, I., Grueschow, M., and Ruff, C. (2014b). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, 82(3):709–720.

R. Anderson, J. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.

Raab, M. (2017). Motor heuristics and embodied choices: how to choose and act. *Current Opinion in Psychology*, 16:34 – 37.

Sarpeshkar, R. (1998). Analog versus digital: Extrapolating from electronics to neurobiology. *Neural computation*, 10:1601–38.

Schieber, M. H. (1990). How might the motor cortex individuate movements? *Trends in Neurosciences*, 13(11):440–445.

Schuman, C. D., Potok, T. E., Patton, R. M., Birdwell, J. D., Dean, M. E., Rose, G. S., and Plank, J. S. (2017). A survey of neuromorphic computing and neural networks in hardware. *ArXiv*, abs/1705.06963.

Shagrir, O. (2010). Brains as analog-model computers. *Studies in History and Philosophy of Science Part A*, 41(3):271–279.

Shagrir, O. (2018). The brain as an input–output model of the world. *Minds and Machines*, 28(1):53–75.

Shen, S. and Ma, W. J. (2016). A detailed comparison of optimality and simplicity in perceptual decision making. *Psychological Review*, 123(4):452–480.

Simon, H. A. and Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2):145–159.

Snider, J., Lee, D., Poizner, H., and Gepshtein, S. (2015). Prospective optimization with limited resources. *PLOS Computational Biology*, 11(9):e1004501.

Söllner, A., Bröder, A., Glöckner, A., and Betsch, T. (2014). Single-process versus multiple-strategy models of decision making: Evidence from an information intrusion paradigm. *Acta Psychologica*, 146:84–96.

Sprenger, A., Dougherty, M., Atkins, S., Franco-Watkins, A., Thomas, R., Lange, N., and Abbs, B. (2011). Implications of cognitive load for hypothesis generation and probability judgment. *Frontiers in Psychology*, 2:129.

Spruston, N. and Johnston, D. (1992). Perforated patch-clamp analysis of the passive membrane properties of three classes of hippocampal neurons. *Journal of Neurophysiology*, 67(3):508–529.

Stiefel, K., Tapson, J., and van Schaik, A. (2012). Temporal order detection and coding in nervous systems. *Neural computation*, 25.

Strait, C., Blanchard, T., and Hayden, B. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron*, 82(6):1357–1366.

Strukov, B., D., Snider S., G., Stewart, D., and Williams, S. (2008). The missing memristor found. *Nature*, 453:80–3.

Summerfield, C. and Tsetsos, K. (2015). Do humans make good decisions? *Trends in Cognitive Sciences*, 19(1):27–34.

Thébault, K. (2019). What can we learn from analogue experiments? In Dardashti, R., Dawid, R., and Thébault, K., editors, *Why Trust a Theory?: Epistemology of Fundamental Physics*, page 184–201. Cambridge University Press.

Todd, M., P. and Brighton, H. (2015). Building the theory of ecological rationality. *Minds and Machines*, 26(1-2):9–30.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., and Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, 113(11):3102–3107.

Turing, A. (1936). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(1):230–265.

Turing, A. M. (1992). Lecture to the london mathematical society on 20 february 1947. In Ince, D. C., editor, *Mechanical Intelligence: Collected Works of A.M. Turing*, pages 87–106. North-Holland Publishing Co.

Ullman, S. (2019). Using neuroscience to develop artificial intelligence. *Science*, 363(6428):692–693.

Ulmann, B. (2013). *Analog Computing*. Oldenbourg Wissenschaftsverlag.

Ulmann, B. (2017). *Analog Computer Programming*. CreateSpace Independent Publishing Platform.

Utagawa, A., Asai, T., and Amemiya, Y. (2011). High-fidelity pulse density modulation in neuromorphic electric circuits utilizing natural heterogeneity. *Nonlinear Theory and Its Applications, IEICE*, 2:218–225.

Utagawa, A., Asai, T., Hirose, T., and Amemiya, Y. (2007). An inhibitory neural-network circuit exhibiting noise shaping with subthreshold mos neuron circuits. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, E90-A(10):2108–2115.

Utagawa, A., Asai, T., Hirose, T., and Amemiya, Y. (2008). Noise-induced synchronization among sub-rf cmos analog oscillators for skew-free clock distribution. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E91.A(9):2475–2481.

von Neumann, J. (1951). The general an logical theory of automata. *Cerebral mechanisms in behavior*, 5:1–41.

Štukelj, G. (2019). On the simplicity of simple heuristics. *Adaptive Behavior*. https://doi.org/10.1177/1059712319861589.

Walter, F., Röhrbein, F., and Knoll, A. (2015). Neuromorphic implementations of neurobiological learning algorithms for spiking neural networks. *Neural Networks*, 72:152 – 167. Neurobiologically Inspired Robotics: Enhanced Autonomy through Neuromorphic Cognition.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, 90(3):1195–1268. PMID: 20664082.

Wegener, I. (1987). *The Complexity of Boolean Functions*. John Wiley & Sons, Inc., New York, NY, USA.

Wong, H.-S. and Salahuddin, S. (2015). Memory leads the way to better computing. *Nature nanotechnology*, 10:191–4.

Wu, S.-W., Delgado, M., and Maloney, L. (2015). Motor decision-making. In *Brain Mapping*, pages 417–427. Elsevier.

Wu, S.-W., Delgado, M. R., and Maloney, L. T. (2011). The neural correlates of subjective utility of monetary outcome and probability weight in economic and in motor decision under risk. *Journal of Neuroscience*, 31(24):8822–8831.

Zhang, D. P., Jayasena, N., Lyashevsky, A., Greathouse, J., Meswani, M., Nutter, M., and Ignatowski, M. (2013). A new perspective on processing-in-memory architecture design. In *Proceedings of the ACM SIGPLAN Workshop on Memory Systems Performance and Correctness*, MSPC '13, New York, NY, USA. Association for Computing Machinery.

Zhang, H. and Maloney, L. T. (2012a). Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, 6:1.

Zhang, H. and Maloney, L. T. (2012b). Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, 6:1.

Zhu, X., Li, D., Liang, X., and Lu, W. (2019). Ionic modulation and ionic coupling effects in mos2 devices for neuromorphic computing. *Nature Materials*, 18.

Curriculum Vitae

Completed Education

M.A. Philosophy, Faculty of Arts, University of Ljubljana, 2014 - 2016.

Exchange student, *Munich Center for Mathematical Philosophy* (MCMP), LMU, 2015/2016.

Hilbert Bernays summer school on Logic and Computation, Faculty of Mathematics and Computer Science, Georg-August-Universität, Göttingen, 2015.

Epicenter course on Epistemic Game Theory, EpiCenter, Maastricht University, 2015.

B.A. Philosophy, Faculty of Arts, University of Ljubljana, 2011 - 2014.

Teaching experience

Teaching assistant, "*Central topics in philosophy of science*" (probability theory, Bayesian networks), J. Landes, MCMP/LMU, Winter 2019/20.

Teaching assistant, "*Formal Methods II - Models and Simulations*" (introduction to agent-based modeling & programming), R. Stewart, MCMP/LMU, Summer 2018.

Teaching assistant, "*Central topics in philosophy of science*" (probability theory, Bayesian networks), S. Hartmann, MCMP/LMU, Winter 2017/18.

Talks & lectures

"On analog neural computation", *MCMP-Western Ontario Workshop on Computation in Scientific Theory and Practice*, Munich Center for Mathematical Philosophy (MCMP), Munich, June 2019.

"Neural networks as an architectural principle", *SAS Workshop 2018*, High-Performance Computing Center (HLRS), Stuttgart, November 2018.

"Probability with Bayesian networks in Python", *Tutorial lecture*, PyData Munich, January 2018.

"Urn models, combinatorics, and the Dutch-book theorem", *Tutorial lecture*, MCMP, October 2017.

"Quine on modal logic", *Second Belgrade Graduate Conference in Philosophy and Logic*, University of Belgrade, 2015.

List of publications

1. ***Štukelj, G. (2019). On the simplicity of simple heuristics. Adaptive Behavior. https://doi.org/10.1177/1059712319861589***

Eidesstattliche Versicherung / Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation "Significance of neural noise" selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation "Significance of neural noise" is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, den 20.04.2020

Munich, 20.04.2020

Gašper Štukelj

Declaration of author contributions

1. Štukelj, G. (2019). On the simplicity of simple heuristics. [Chapter 2] Adaptive Behavior. doi: 10.1177/1059712319861589

The author of the thesis, Gašper Štukelj, is the first and sole author of the publication.

2. Štukelj, G. (unpublished). On analog neural computation. [Chapter 3]

The author of the thesis, Gašper Štukelj, is the first and sole author of the manuscript.

3. Štukelj, G. (unpublished). On brain analogs: The curious case of neuromorphic electronics. [Chapter 4]

The author of the thesis, Gašper Štukelj, is the first and sole author of the manuscript.

München, den 20.04.2020 Gašper Štukelj Prof. Dr. Stephan Hartmann