



Computational indeterminacy and explanations in cognitive science

Philippos Papayannopoulos¹ · Nir Fresco² · Oron Shagrir³

Received: 14 February 2022 / Accepted: 2 September 2022 / Published online: 7 November 2022
© The Author(s) 2022

Abstract

Computational physical systems may exhibit indeterminacy of computation (IC). Their identified physical dynamics may not suffice to select a unique computational profile. We consider this phenomenon from the point of view of cognitive science and examine how computational profiles of cognitive systems are identified and justified in practice, in the light of IC. To that end, we look at the literature on the underdetermination of theory by evidence and argue that the same devices that can be successfully employed to confirm physical hypotheses can also be used to rationally single out computational profiles, notwithstanding IC.

Keywords Indeterminacy of computation · Underdetermination of theory by evidence · Computational explanations · Computational individuation

Introduction

Studying physical systems under the assumption that they compute functions or implement specific automata has been a fruitful practice in various areas of scientific research—from the physics of information to computational neuroscience. Nevertheless, characterizing a system computationally is generally a challenging task,

✉ Philippos Papayannopoulos
fpapagia@uwo.ca

Nir Fresco
nfresco@bgu.ac.il

Oron Shagrir
oron.shagrir@gmail.com

¹ Department of Philosophy, The Hebrew University of Jerusalem, 91905 Mt. Scopus, Jerusalem, Israel

² Departments of Cognitive and Brain Sciences & Philosophy, Ben-Gurion University of the Negev, 8410501 Beer Sheva, Israel

³ Departments of Philosophy & Cognitive and Brain Sciences, The Hebrew University of Jerusalem, 91905 Mt. Scopus, Jerusalem, Israel

even when the system's physical behavior is sufficiently known. As a simple example, consider a physical system whose input and output behavior can be described as implementing an OR-gate; that is, certain physical states have been mapped to the logical values True/False (computational states 1/0, respectively) in accordance with Table 1. By swapping every logical (computational) state that is mapped to a given physical state for its inverse—that is, by assigning 0 to any physical state that was previously mapped to 1, and vice-versa—the same system can now be described as implementing an AND-gate (Table 2).

What determines whether the specific system, in this example, implements disjunction or conjunction is how the inputs and outputs are interpreted (i.e., as a computational state True or False). This kind of indeterminacy may even extend beyond Boolean functions to functions in discrete and continuous mathematics, such as addition and multiplication over the real numbers.¹ And yet, this is not the only way in which the same physical behavior may support multiple computations simultaneously. As Shagrir (2001) and Shagrir (2012) show, some physical systems can be seen as implementing multiple finite-state automata concurrently (below, we present such an example). The upshot is that there exist physical systems that can be seen as either simultaneously computing more than one mathematical function, or simultaneously implementing more than one formal structure (or even both). Let us refer to this phenomenon as 'Indeterminacy of Computation' (IC).²

This multiplicity of computational profiles notwithstanding, computational explanation in scientific practice picks out and relies on only one of these possibilities. But the IC challenge concerns exactly that: it is not always possible to select a unique implemented automaton or computed function as *the* computational identity of the system based on its physical properties alone.

Hence, the question arises: given the IC challenges imposed on computational individuation, what justifies computational explanations that rely on singling out specific computational profiles over others? How do cognitive scientists decide between the simultaneously implemented structures—or the different computed functions—in explaining a system's observed success in some cognitive task, and on what basis can the advanced explanations be considered well-determined and warranted?

We propose that valuable insight into these questions may be gained from extant responses to another case of non-determinacy: the *underdetermination of theory by evidence* (UTE). Conjecturing that a cognitive system computes a specific function can be seen as a special case of hypothesis testing. Thus, the challenge posed by UTE seems, *prima facie*, also pertinent to the way that a cognitive scientist chooses between rival hypotheses. A large body of literature has sought to tackle UTE by formulating accounts of confirmation that go beyond naïve hypothetico-deductivism (e.g., based on theoretical virtues of theories, statistical methods, Bayesian reasoning, etc.). In this article, we make the case that similar arguments can also be

¹ See Fresco et al. (2021) and Papayannopoulos et al. (2022) for examples.

² The term is due to Jack Copeland.

brought to bear on the IC challenge in relation to computational hypotheses in cognitive science.

More specifically, the thesis we put forward is that in studying neuronal computations, we can be aided in selecting computational profiles (thereby overcoming IC concerns) on similar grounds that are used to choose between rival hypotheses in physical science (in overcoming UTE). In other words, the same tools that help us justifiably rule out specific hypotheses as less plausible for a given phenomenon also help us to rule out *computational hypotheses* as non pertinent to neuronal computations within given contexts.

Section “[The IC in more detail](#)” discusses IC in more detail and provides some additional examples. Section “[The current debate on computational individuation and the relation between IC and UTE](#)” briefly reviews the current debate about computational individuation. Section “[The Underdetermination of Theory by Evidence](#)” concerns UTE: although the thesis allegedly poses a special kind of challenge to confirmation of theories (Sect. [The UTE thesis](#)) there is a variety of conceptual and mathematical devices that provide us with grounds for deeming certain hypotheses preferred and better confirmed than others (Sect. [Responses to UTE](#)). Section “[The confirmation methods at work: determining computational hypotheses](#)” presents a detailed case study from computational neuroscience and argues that the same devices can also be used to confirm hypotheses within a computational context; that is, hypotheses concerning computational profiles, algorithms, and their implementations. Section “[Determining the indeterminate? The confirmation methods against IC](#)” shows how the same, aforementioned devices can be also appropriate for systems that admit simultaneously of several ways of grouping physical and computational states (IC). Finally, section “[Conclusions](#)” provides some conclusions and ties up loose ends.

The IC in more detail

Let us make some general observations about IC and see some more examples. In what follows, we will adopt a distinction proposed by Papayannopoulos et al. (2022) between two kinds of IC; an *interpretative* and a *functional* kind. Both species of indeterminacy pertain to computational individuation to a greater or lesser extent. Roughly, *interpretative* IC concerns an indeterminacy regarding the identification of the logical or mathematical function that the system computes; such indeterminacy arises owing to the possibility of more than one way of assigning logical and/or mathematical content to specific physical (abstract) states. The example from the previous section is such a case, since it is indeterminate whether the system computes an OR or an AND function. *Functional* IC concerns an indeterminacy regarding the proper way of grouping together physical properties (or states) in order to determine the functional organization of the computing system. The next example falls under functional IC, because the indeterminacy between the two obtained computational profiles in Tables 4 and 5 arises by virtue of variant partitions of the same set physical states (i.e., by virtue of more than one possible functional organization).

Table 1 The logical input-output relations of an OR-gate

Input A	Input B	Output
1	1	1
1	0	1
0	1	1
0	0	0

Table 2 The logical input-output relations of an AND-gate

Input A	Input B	Output
0	0	0
0	1	0
1	0	0
1	1	1

Consider a physical system \mathcal{P} with two voltage inputs and one voltage output, as in Fig. 1. The system can be thought of as a kind of a black box, whose computational behavior we are trying to identify. Suppose that after an appropriate number of carefully designed measurements, we end up with a taxonomy of the input/output dynamics depicted in Table 3; that is, we have come to specify three voltage ranges—Low (L), Medium (M), and High (H)—as computationally stable.

Given a dynamic behavior of the system as in Table 3, how should we characterize it computationally? Without loss of generality, let us assume that our working hypothesis is that the system computes some Boolean function. One possible way is to map the High range to some logical state (say, T) and the other two, Medium and Low grouped together, to its opposite (i.e., F). Under this mapping, the system computes an XOR function (as in Table 4). But an equally plausible route is to group together the High and Medium ranges, assigned to T, and map Low alone to F. Under that second mapping, the system computes an OR function (as in Table 5). This is a case of indeterminacy of the functional profile of the system, consequently giving rise to functional IC.³

Before closing this section, let us consider a final example, which will be also useful in our later discussions about possible indeterminacies arising in our case study. Assume a physical system \mathcal{Q} that is again a tri-stable flip-detector, like the system \mathcal{P} above, and suppose that by means of appropriate measurements we obtain the input/output profile in Table 6.

One way to characterize \mathcal{Q} computationally is to determine some functional organization by abstracting from the physical voltages. For example, by mapping

³ Note that the same system exhibits also interpretative IC, in the sense that the truth-values in Tables 4 and 5 can be swapped for their inverse, thus leading to an XNOR- and an AND-gate, respectively.

Fig. 1 The physical system \mathcal{P} with three stable states

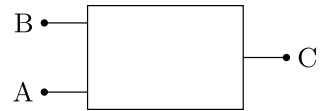


Table 3 A taxonomy of the input/output behavior of the system \mathcal{P} . The system emits 7–10 volts, if it receives a signal greater than 7 volts from exactly one input channel. It emits 0–3 volts, if it receives less than 3 volts from both input channels, and it emits 4–6 volts in any other case

Input A (V)	Input B (V)	Output C (V)
7–10 (H)	7–10 (H)	4–6 (M)
7–10 (H)	4–6 (M)	7–10 (H)
7–10 (H)	0–3 (L)	7–10 (H)
4–6 (M)	7–10 (H)	7–10 (H)
4–6 (M)	4–6 (M)	4–6 (M)
4–6 (M)	0–3 (L)	4–6 (M)
0–3 (L)	7–10 (H)	7–10 (H)
0–3 (L)	4–6 (M)	4–6 (M)
0–3 (L)	0–3 (L)	0–3 (L)

Table 4 The system implements an XOR-gate, when M and L are both assigned the truth-value F

Input A	Input B	Output C
T	T	F
T	F	T
F	T	T
F	F	F

Table 5 The system implements an OR-gate, when H and M are both assigned the truth-value T

Input A	Input B	Output C
T	T	T
T	F	T
F	T	T
F	F	F

the higher voltage ranges, 0.5–0.9V, to the logical state T and the lower and medium ranges, 0–0.5V, to F, we obtain an AND gate as the computational identity of the system; see Table 7.⁴ But now observe that the mapping relating the physical inputs and outputs in Table 6 is also consistent with a multiplication description:

$$V_{\text{out}} = V_1 \cdot V_2$$

⁴ The reverse mapping (0–0.5V) → T and (0.5–0.9V) → F, gives an OR; thus the system is also interpretative indeterminate.

Table 6 A taxonomy of the input/output behavior of the system \mathcal{Q}

Input A (V)	Input B (V)	Output C (V)
0.8–0.9	0.8–0.9	0.64–0.81
0.8–0.9	0.4–0.5	0.32–0.45
0.8–0.9	0–0.2	0–0.18
0.4–0.5	0.8–0.9	0.32–0.45
0.4–0.5	0.4–0.5	0.16–0.25
0.4–0.5	0–0.2	0–0.1
0–0.2	0.8–0.9	0–0.18
0–0.2	0.4–0.5	0–0.1
0–0.2	0–0.2	0–0.04

Table 7 The system \mathcal{Q} implements an AND-gate, when values between 0.5–0.9V are grouped together under the truth-value T

Input A	Input B	Output C
T	T	T
T	F	F
F	T	F
F	F	F

This, then, gives rise to an alternative abstract/functional structure, according to which \mathcal{Q} computes a multiplication function:

$$\text{Output}_C = \text{Input}_A \cdot \text{Input}_B \tag{1}$$

This is again a case of functional indeterminacy, because it arises on account of different possible functional organizations. But it is interesting to note that the functional indeterminacy of this example stems from groupings of physical properties at different levels of granularity. The functional profile of a multiplication operation arises from a more fine-grained carving of the physical state space than that of the AND/OR operation. On the other hand, the two different functional profiles of Tables 4 and 5 stem from variant groupings that exist at the same level of granularity.

The current debate on computational individuation and the relation between IC and UTE

The debate

An ongoing debate concerns the nature of computational characterization itself; that is, what it means to computationally identify a system like the one above —what its computational profile should look like. Participants in the debate aim to formulate

appropriate criteria, on the basis of which a unique computational profile can always be singled out. Nonetheless, there is no agreement on what such criteria should be like.

A main dividing line has to do with whether the critical parameter that singles out the computational identity from all the simultaneous implementations/computations is *intrinsic* or *extrinsic* to the system. Some proponents of the intrinsic view suggest that the computational profile of the system (which also captures its computing mechanism) should be identified with a (unique) abstract structure that underpins a stable contribution to the goals of the system. Other possible abstract structures in the system do not have the same contribution.⁵ For example, Dewhurst (2018) proposes that the computational profile of a system like \mathcal{P} above is to be provided by Table 3, instead of some table describing computational functions, such as XOR (Table 4) or OR (Table 5). Coelho Mollo (2018) propounds that the computational profile should be identified with the functional profile of the computing system, which for our toy example would be given by a Table like 8.

Proponents of the *extrinsic* view argue that computational individuation requires taking into account parameters that are *external* to the computing system. Piccinini (2015) suggests that the computational profile can vary across contexts; it depends on the kind of interaction between the system and its close environment. Fresco and Miłkowski (2021) and Fresco (2021) agree with Piccinini that computation is context-dependent, but offer a different account of the nature of the system-environment interaction. Pragmatists suggest that the feature that singles out the relevant computation depends on the explanatory aims of the scientists (Egan 2012; Rescorla 2014; Matthews and Dresner 2017). Proponents of the semantic view of computation suggest that the missing ingredient is the content of the physical states (e.g., Sprevak 2010). Shagrir (2001, 2020), who also adopts a semantic approach, further distinguishes between the notions of ‘implementation’ and ‘computation’: implementation is non-semantic whereas computation is content-dependent.

The stance in this paper

The big interest of the above proposals notwithstanding, in this article we look into the issue from a different perspective. Rather than focusing on an *ontic* understanding of computational individuation in the light of IC, we want to draw attention to its *epistemic* status. That is, instead of tackling the question of what it is to computationally characterize a system —what are the relevant causal, modal, functional, mechanistic, semantic parameters and whatnot— we ask the question: how does the goal of computationally individuating a system get achieved within scientific practice? How are specific explanatory computational hypotheses discovered and identified in practice, and how are they justified in the light of the IC challenge?

⁵ The goal of the system is often seen in terms of what the system was designed to do or has been used for. The latter are further characterized in evolutionary (Millikan 1984), learnability (Dretske 1988), dispositional (Maley et al. 2047), or other terms.

The literature proposals presented above do not suffice to illuminate this epistemic question, since no appeal to modal, structural, mechanistic, functional, semantic or other philosophical constraints seem to explicitly play a role in the process of scientists' computationally individuating a system in published research. On what other grounds, then, can cognitive scientists be justified when they pin down specific computations as the ones appropriate to explain the success of their studied systems instead of others?

We attempt to answer this question by drawing on lessons learned from philosophy of science in its dealing with the UTE challenge. Extensive work has tried to identify methods and reasons based on which scientists are justified in deeming specific hypotheses preferable to others, given a body of observational data. We argue that similar considerations justify cognitive scientists' computational hypotheses too.

Cognitive scientists do so in two ways. First, researchers try to identify mathematical descriptions that fit with the experimental data; concerned in our context with neuronal firing rates, recordings of membrane potentials, etc. Importantly, this is already one instance of hypothesis confirmation, based on the methods identified by the UTE and confirmation literature. In comparison with the simple example from above, it can be seen as the equivalent of identifying the *dynamical* profile summarized in Table 3 (or Table 6). But this, considered in and of itself, does not suffice to uniquely determine a *computational* profile for the system; different interpretations and/or groupings of physical (or functional) states can give rise to different computational profiles (viz., interpretative and/or functional IC). The second way in which the methods identified by the UTE literature help, then, is in determining interpretations and groupings that are better suited for explaining the studied system's behavior. Here, the role of context (environment in which the system is embedded) becomes crucial.

IC vs. UTE and the role of context

We have just said—and will argue in detail in Sect. [Determining the indeterminate? The confirmation methods against IC](#)—that context is crucial for identifying a system's computational profile, even when its dynamical behavior is already known. But how does this square with ontic intrinsic views on computational individuation, which take a system's computational profile to be some *fixed*, context-independent structure? Answering this question will clarify our presupposition that the discovery of computational individuation is not necessarily entangled with its ontic description, so it would be instructive to do so. As an example, recall Coelho Mollo's (2018) view, according to which the fixed computational structure of system \mathcal{P} (Fig. 1) is given by Table 8—meaning that there is no IC problem for the scientists to overcome while individuating the system. Now, even if we adopt this view, it is still the case—and universally agreed—that the system also implements (or can be mapped to) the simpler formal structures XOR and OR from Tables 4 and 5. The difference is that, on this view, these simpler structures just are not part of the *computational* structure (profile) itself. Our claim is that even if Coelho Mollo (2018)

Table 8 The functional profile of system \mathcal{P} (A, B, and C effectively correspond to H, M and L). This structure implies an XOR-gate, if A is mapped to computational state T and B, C together to F, and implies an OR-gate, if A, B are mapped together to T and C to F (Tables 4 and 5)

Input A	Input B	Output C
A	A	B
A	B	A
A	C	A
B	A	A
B	B	B
B	C	B
C	A	A
C	B	B
C	C	C

is right about computational individuation —e.g., even if his hypothesis about \mathcal{P} 's computational profile being the structure in Table 8 is correct— it is still on account of the confirmation/UTE methods (plus some contextual clues) that scientists would discover and justify such a hypothesis and exclude any rival ones —e.g., about the simpler structures XOR and OR.

Although our aim here is to draw lessons from the philosophical responses to UTE and bring them to bear on the IC challenge, we do not mean to suggest that the two cases of non-determinacy are similar or analogous. On the contrary, the two challenges are very different in nature, as even implied by the terms ‘underdetermination’ and ‘indeterminacy’. Presumably, the ‘underdetermination’ part in ‘UTE’ implies that although evidence never determines a unique theory, there does exist some relevant fact of the matter; albeit, unreachable by us. On the other hand, the ‘indeterminacy’ part in ‘IC’ implies that no fact of the matter exists either; the system \mathcal{P} (Table 3), for example, is computationally indeterminate in the sense that no fact of the matter prescribes whether its computational profile is the XOR-gate or the OR-gate (or some other profile).⁶

These differences between UTE and IC notwithstanding, there is a clear sense in which both cases pose the challenge of justifiably giving preference to specific hypotheses over others, in spite of some possible inherent non-determinacy. This is what motivates our proposal that responses to UTE—that is, reasons for considering certain hypotheses more justified— can be brought to bear also on deeming hypotheses about computations more justified than others. But how can such a proposal be sensibly propounded, given our foregoing remark that, contrary to UTE, IC indicates that no actual fact of the matter exists?

The key idea we rely on, for this answer, is ‘context’. Recall that we are here concerned with these questions from the point of view of a scientist who examines

⁶ Supporters of the (ontic) intrinsic view of individuation may deny this statement, if they read it metaphysically. Nevertheless, note that all the talk about ‘fact of the matter’ here is meant to be read *epistemically*; i.e., as referring to solid epistemic—and not necessarily metaphysical— grounds.

systems that already operate and fulfill specific functions within certain environments and as parts of greater wholes. That means that we can reasonably assume that investigating the system as part of its environment plays an essential (epistemic) role, enabling us to restrict the number of plausible groupings and interpretations.⁷

Therefore, although a physical system may be computationally indeterminate *simpliciter*, in the sense that no relevant fact of the matter exists, we suggest that it is not as such while considered operating within a particular context, and that a relevant fact of the matter does exist. Note, however, that the claims made here are to be read solely from an epistemic standpoint, and not as metaphysical assertions whatsoever.

The underdetermination of theory by evidence

In this section, we discuss UTE as a preamble for assessing whether methods applied in response to UTE may be brought to bear when competing —yet equally privileged— computational hypotheses are concerned in the context of IC. In slogan form, UTE asserts: “No body of data or evidence or observation can determine a scientific theory or hypothesis within a theory” (Norton 2008). That said, there are different varieties of the thesis in the philosophical market. We discuss first what sense of UTE concerns us here.

The UTE thesis

We should clarify from the outset that UTE is not just the (undeniable) claim that it *so happens, sometimes*, that the available body of data does not suffice to uniquely determine a suitable explanatory hypothesis. It is the much stronger thesis that in *every case* and for *any amount* and *any kind* of data —no matter how large the amount or how ingenious its collection methods may be— there is no hypothesis that can be singled out as preferred and better confirmed than the rest. And, additionally, UTE is not just the claim that some parts of a theory may happen to be underdetermined by all experimental evidence, which is also undeniably true.⁸ Rather than concerning certain parts of some theories, the underdetermination threat supposedly hangs over any theory or hypothesis.

What can be safely inferred about UTE, given its really bold scope? Suppose that we have some evidence E , and that we formulate an explanatory hypothesis H_1 for it. By UTE, H_1 cannot be singled out as preferred to other alternative hypotheses. Therefore, there must be at least some H_2 , such that H_2 is *equally* confirmed by E ; for, otherwise, the one of the two hypotheses that is deemed better confirmed would

⁷ See, e.g., Harbecke and Shagrir (2019) for a discussion on the role of the context in computational explanations.

⁸ For example, in Newtonian mechanics, we can have observationally equivalent pairs of theories by considering different absolute states of rest (Norton 2008).

also be preferable. Thus, for any body of data, there must always exist at least two, equally confirmed explanatory hypotheses.

As Norton (2008) points out, however, empirically equivalent pairs of theories (or hypotheses) can legitimately be taken as equally confirmed “only if all that matters in the confirmation relation is that the two theories have identical observational consequences.” But, as he also points out, this, in turn, is no different from naïve hypothetico-deductive (HD) confirmation. That is, the idea that a hypothesis H , conjoined with auxiliary assumptions, is confirmed by evidence E , insofar as H logically entails E .⁹ Nevertheless, HD is widely regarded as a poor account of confirmation, since, on the one hand, it may be unacceptably permissive (e.g., the problem of vacuous conjunctions) and, on the other hand, it may not capture cases of evidence that we would consider confirmatory even though it may not be logically entailed by the hypotheses.¹⁰ Indeed, the literature in confirmation has been at pains to show how ampliative inference in scientific practice involves substantially more than just HD.¹¹

Responses to UTE

Let us now assess the main scientific methods that are identified in the confirmation literature as strong responses to UTE, in order to later examine their applicability in the case of competing computational hypotheses arising from IC. What other standards of empirical warrant for certain hypotheses, besides logical entailment of the observed phenomena, do exist out there? Norton (2008) offers a helpful taxonomy of guiding principles, as they appear in the literature. Briefly reviewing it will be useful for the upcoming discussion of our case study.

A first family of approaches is grounded in the assumption that the relation between evidence and hypotheses can be represented and formalized on the basis of the calculus of probabilities. The dominant approach is ‘Bayesianism’, with the theorem under the same name at its heart. A second family includes accounts and strategies of confirmation that augment naïve HD with additional requirements in order to restrict its undesirable permissiveness. Prominent approaches emphasize various theoretical virtues, like simplicity, explanatory power and unifying power of certain hypotheses as grounds for additional confirmation of them.¹²

An important group of approaches includes what Norton (2008) calls ‘exclusionary accounts’. According to them, evidence E confirms a hypothesis H to the extent that H entails E but *also* E would be very unlikely, were H false. This principle underlies *statistical hypothesis testing* as well as *controlled group* studies (e.g., for a new medical treatment), and it is also prevalent in our case study.

⁹ See, e.g., (2.2 Salmon et al. 1992; Crupi 2016).

¹⁰ For some standard examples and detailed discussions on UTE and HD, see, e.g., (Laudan 1990; Laudan and Leplin 1991; Norton 2008).

¹¹ An inference is ampliative if the conclusion has content that goes beyond the content in the premises (for example, induction). See, e.g., (sec. 1.4 Salmon et al. 1992).

¹² See, e.g., (ch.8 Psillos 1999) for some arguments to the effect that theoretical virtues like these effectively undermine UTE.

In this paper, we focus on a very interesting, yet less known, method, called ‘demonstrative induction’ (DI) (aka ‘Newtonian deduction from the phenomena’). It consists in *deducing* a target theory, or hypothesis, from the evidence, with the assistance of some general auxiliary hypotheses. Since the inference is deductive, there is no longer an inductive risk in the inferential steps. The risk is relocated into accepting the more general auxiliary hypotheses. Importantly, the stronger and more independent the reasons we have to accept the general auxiliary hypotheses, the more secure the inferred target hypothesis is. Furthermore, since the inference is from evidence to hypothesis, UTE is clearly undermined; for, in that case, evidence does point to a unique theory or hypothesis. The method, however, is not always easy to implement, as it has generally a narrow scope of applicability. Yet, it can be found underlying some crucial moments of theory development in the history of science.¹³ We will say more about this method, and see how it fits with our case study, in Sect. [How computational hypotheses are confirmed](#).

The confirmation methods at work: determining computational hypotheses

A neuroscientific case study

We examine a representative case of characterizing computationally the responses of a visual neuronal system in the locust’s brain to approaching objects on a collision course. This particular case offers a convenient system to be used as a model to study how a single neuron and its presynaptic network may implement mathematical operations.

Vision is crucial in notifying animals of imminent dangers, such as an impending collision with a predator or a surface. Two motion-sensitive neurons have been thoroughly investigated in connection with that, owing to their believed involvement in the generation of escape behaviors in response to looming stimuli. First, the response of the descending contralateral motion detector neuron (DCMD) to approaching and translating objects was initially investigated in Hatsopoulos et al. (1995). The DCMD relays spikes in a 1:1 manner to thoracic motor centers (Gabbiani et al. 2002). Synapsed onto DCMD is another large neuron, the lobula giant movement detector (LGMD). The LGMD receives various synaptic inputs—most notably a feedforward excitation and two distinct feedforward inhibitions (Fig. 2)—and it is where the studied computation is thought to be carried out. The connection between the LGMD and

¹³ DI has been largely neglected in the philosophical literature of the early 20th century but has been recently rediscovered and discussed by a number of philosophers of science. It has been shown that it was employed by Newton to deduce his inverse square law of gravitation (Dorling 1990; Harper 1990, 1997; Norton 2000), by Einstein in his discovery of general relativity (e.g., DiSalle et al. 1994; Dorling 1995; Norton 1995), by Bohr to deduce the quantization of energy levels (Norton 2000), and by a number of other physicists as well. For a variety of examples of the use of DI in milestone events in the history of physics, as well as for a formalization of the method as an inferential logical schema, see Dorling (1973).

DCMD is so strong that each action potential in the LGMD elicits an action potential in the DCMD, and, conversely, each action potential in the DCMD is caused by an action potential in the LGMD (Gabbiani et al. 1999, 1122).

The proposed hypothesis for the computation of the LGMD is that it computes the time in which an approaching object on a collision course with the animal reaches a constant angular size in the animal’s retina (e.g., Hatsopoulos et al. 1995; Gabbiani et al. 1999). To test the hypothesis, various locusts, in a series of controlled experiments, were exposed to simulated approaching objects. The visual stimuli were generated on computer monitors, by means of simulated dark squares of various sizes approaching with constant velocity and on a collision course with the animals. The responses of the LGMD and DCMD to looming objects were found to be as follows. The firing rate starts early during the approach phase and then gradually increases as the object grows larger, as if these cells are “tracking” the object over its approach. Then it peaks and eventually decreases (Gabbiani et al. 1999, 1125). The peak firing rate occurs before collision but, for each animal, always a fixed delay, δ , after the object has reached a certain angular size, θ_{thr} , on the retina; hence, the proposed hypothesis that the LGMD computes the time that θ_{thr} is reached.

How is this computation realized? To formulate an answer to this question, a mathematical characterization of the firing response of the LGMD is first sought; one that fits the experimentally obtained firing profile, depicted in Fig. 3. The mathematical description that fits the experimental points, determined by the researchers, has the general form (Gabbiani et al. 2002):

$$f(t) = g(\psi(t - \delta) \cdot e^{-a\theta(t-\delta)}) \tag{2}$$

where $\psi(t) = \frac{\dot{\theta}(t)}{2}$, $a = (\tan \frac{\theta_{thr}}{2})^{-1}$, and g some time-independent non-linear parameter. We now explain the terms in Eq. (2), which captures a class of functions, in accordance with different circumstances of object approaching.

The variable that characterizes the object’s approach is the time course of the angular size θ subtended by the object on the locust’s retina (Fig. 4). Let the object’s distance from the eye be x ($x = 0$ means collision), the time before collision be $t < 0$ ($t = 0$ at collision), and the object’s velocity be v . Then, for constant approaching velocity $v < 0$:

$$x(t) = vt$$

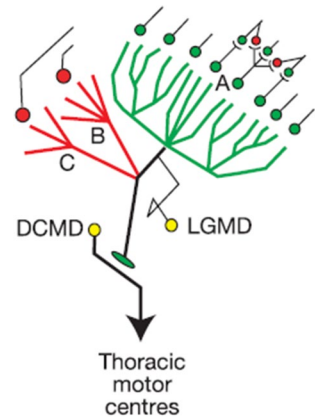
From trigonometry (Fig. 4):

$$\tan \frac{\theta(t)}{2} = \frac{l}{x(t)} = \frac{l_{scr}(t)}{x_{scr}}$$

and so:

$$\theta(t) = 2 \tan^{-1} \frac{l}{vt} \tag{3}$$

Fig. 2 Schematic of the LGMD neuron synaptic input. The green dots (fan A on LGMD) represent received feed-forward excitation, while the red dots (fans B and C) represent received feedforward inhibition. The smaller red dots represent lateral inhibition, presynaptic on the excitatory pathway (A). (From Gabbiani et al. (2002), with permission. Synaptic inputs to the DCMD neuron are not depicted.)



where l is the object’s half size, l_{scr} the simulated object’s half size on the screen, and x_{scr} the distance between the screen and the eye. From (3) by differentiation, one obtains for the angular edge velocity of the object:

$$\dot{\theta}(t) = \frac{d\theta}{dt} = -2 \frac{\frac{l}{v}}{t^2 + (\frac{l}{v})^2}$$

Therefore, the related term $\psi(t)$ in Eq. (2) is:

$$\psi(t - \delta) = \frac{\dot{\theta}(t - \delta)}{2} = - \frac{\frac{l}{v}}{(t - \delta)^2 + (\frac{l}{v})^2}$$

Equation (2) captures the experimental data points to a satisfactory extent, as can be seen by the smooth line in Fig. 3 (middle). The term a in (2) is a constant related to the threshold angle, θ_{thr} , subtended by the object in the retina (i.e., $\tan \frac{\theta_{thr}}{2} = \frac{1}{a}$). The time-independent non-linear parameter, g , characterizes the transformation between the kinematic part, $\psi \cdot e^{-a\theta}$, and the firing rate during approach (dependent on the parameter $\frac{l}{v}$).

What computational hypotheses are in play

Let us take stock of the computational hypotheses involved in this study. At the *computational* level,¹⁴ the LGMD is said to compute the time at which the approaching object has reached a certain angle, θ_{thr} , at the locust’s retina. This angular threshold

¹⁴ We adopt a terminology in similar lines with Marr (1982), and distinguish between a *computational*, an *algorithmic*, and an *implementation* level. In doing so, we also remain faithful to the approach taken by Gabbiani et al. (1999). We note, though, that the latter approach is not always in perfect agreement with Marr’s distinctions between the levels. See also fn.15.

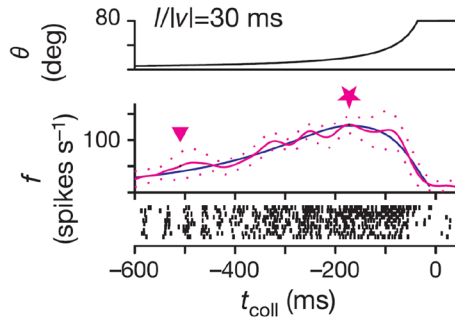
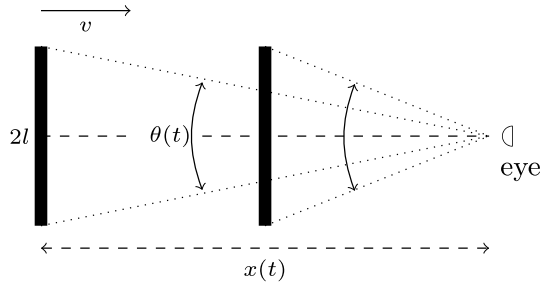


Fig. 3 Responses of the LGMD to approaching squares. Top: angular size of the stimulus as a function of time relative to collision (t_{coll}). Middle: mean instantaneous firing rate \pm s.d. (magenta line and dots). Blue: fit with multiplicative model. Bottom: spike rasters (10 trials). The star represents the peak instantaneous firing rate, while the triangle represents a threshold firing rate of 50 spikes per second. (From Gabbiani et al. (2002), with permission)

Fig. 4 A diagram illustrating the stimulation of the locust's eye by presenting squares of half-size l approaching at a constant velocity v



is reached at a certain time before the peak of the firing rate of the neuron modeled by Eq. (2).

How is the computation carried out at the *algorithmic* level? As we have said, the LGMD receives distinct input projections (Fig. 2), motion-sensitive and size-sensitive, which suggests a multiplication operation implemented by the LGMD, in accordance with Eq. (2). The excitatory retinotopic projection is sensitive to motion, whereas the inhibitory inputs are sensitive to size. Four algorithmic steps can be distinguished: (a) computation of the size subtended by the object at the retina, (b) computation of the angular velocity of the edges of the expanding object on the retina, (c) multiplication of the two, and (d) transformation of the results into a firing rate via the parameter g (Gabbiani et al. 1999).

At the *implementation* level, the hypothesized computation may be carried out in the following ways. One way could be for the two multiplied parameters to be represented directly by the excitatory and inhibitory inputs of the LGMD and then be multiplied. The multiplication could be implemented by means of shunting

inhibition of the velocity signal on the primary neurite (Hatsopoulos et al. 1995). Another way could be as follows: since the excitatory and inhibitory inputs of a neuron are added up, the motion-dependent part could be (presynaptically) represented by excitatory inputs that are logarithmic in angular velocity (i.e., $\log \psi$), and the size-dependent part to be (presynaptically) represented by inhibitory inputs that are proportional to the angular size (i.e., $-a\theta$). Then a postsynaptic summation of both inputs, followed by an approximate exponential transformation of the output, would effectively result in a multiplication operation, by virtue of the equivalence:

$$\psi \cdot e^{-a\theta} = e^{(\log \psi - a\theta)} \quad (4)$$

A third alternative would be for part of the non-linear interaction between the motion-dependent excitation and size-dependent inhibition to partially occur presynaptically, via a lateral inhibitory network that has been identified to protect excitatory synapses onto the LGMD from habituation to whole field motion (Gabbiani et al. 1999, 1139; Gabbiani et al. (2002), 321).

Is a choice between these alternative implementation-level hypotheses underdetermined?¹⁵ Arguably not. They seem amenable to experimental testing; for example, by intervening with the relevant input streams. Based on a following series of experimental tests, then, the researchers put forward the logarithmic-exponential transformation hypothesis as the better confirmed one (i.e., the second of the three options). But the question arises: what grounds exactly are there for having epistemic confidence on such verdicts about the preferred hypotheses at each corresponding level?

How computational hypotheses are confirmed

Let us illustrate how the methods and stratagems used by scientists to determine that specific empirical hypotheses are better *confirmed* than others (*contra* UTE) can also be pertinent to all three levels —computational, algorithmic, and implementation— of computational hypotheses.¹⁶ We do not yet consider, in this section, the possibility of either interpretative or functional IC in our system; at this stage of the argument, we only show how the confirmation methods are also employed to recognize and justify computational claims. In the next section, we will include IC in the picture and argue for our central thesis: that by the very process of forming

¹⁵ The above-discussed hypotheses are not about implementations per se, but rather describe lower-level algorithms. Such algorithms shape the implementation details to a maximum degree, yet technically still belong in the algorithmic level. Although the consideration of them in the context of implementation-related hypotheses (as in Gabbiani et al. 1999, 1139) is not in perfect agreement with Marr's (1982) three-level scheme, we still retain the researchers' taxonomy, since this minor dissonance does not affect our purposes here.

¹⁶ By 'computational hypotheses' we mean mathematical descriptions that can refer either to the input-output function ('computational level'), or to the mediating apparatus in one of two separate ways: a higher-level (i.e., the 'algorithmic level') and the neuronal implementation (i.e., the 'implementation level'). Thus, the term 'computational hypothesis' here is broader than (and includes) the term 'computational profile' (or 'identity').

and justifying computational hypotheses in just the way described in this section, we *automatically* and *simultaneously* pick out the correct computational profile for the given context, even if it were the case that multiple functional and/or computational structures existed simultaneously in the system.

Computational and algorithmic level

At the computational level, concerning *what* is computed, several hypotheses are considered in Hatsopoulos et al. (1995), based on the assumption that the relevant visual neuron monitors properties of the image of the object projected on the retina. Relevant candidate properties are the angular size, θ , and the angular edge velocity, $\dot{\theta}$, which can both be determined monocularly at the retina. But this still can entail different computations, since one can identify more than one function whose output solely relies on quantities determined at the retina. For example, one such function could be the following:

$$\tau(t) = \frac{x(t)}{\dot{x}(t)} = \frac{\sin \theta(t) \cdot \cos \theta(t)}{\dot{\theta}(t)} \approx \frac{\theta(t)}{\dot{\theta}(t)} \text{ (for small } \theta \text{)}$$

and an alternative could be its reciprocal, $\frac{1}{\tau(t)}$. These functions can well be approximated by mere knowledge of θ and $\dot{\theta}$ and could be encoded in the firing rate of the neuron, thereby bringing about an escape command when $\tau(t)$ has decreased below a certain threshold, or, respectively, when $\frac{1}{\tau(t)}$ has exceeded a specific threshold. Nevertheless, additional evidence via controlled experiments show that the time of the peak firing rate is strongly correlated with the collision time, and the delay between peak firing rate and collision depends on both l (object size) and v (object velocity). Since $\tau(t)$ is independent of these parameters, this gives extra reasons for rejecting the hypothesis that the LGMD encodes either $\tau(t)$ or $\frac{1}{\tau(t)}$ (Hatsopoulos et al. (1995), 1000-1).

Hatsopoulos et al. (1995), then, offer strong reasons to regard the computational hypothesis modeled by Eq. (2) as better confirmed by the relevant data than the others. Such strong reasons are grounded in the same rules of ampliative inference that render naïve HD and UTE unjustifiable; namely, carefully designed controlled experiments and tests. But, in a following paper (Gabbiani et al. 1999), the researchers offer an even stronger argument for (2). In fact, we submit that their argument in that paper can be seen as a clear instance of *demonstrative induction* (DI). Let us explain what we mean by that. We will first digress a little to give an example of DI from the history of physics (taken from Norton 2000), and then will explain how the argument in Gabbiani et al. (1999) can be seen as an instance of the same method.

The method, recall, (Sect. [Responses to UTE](#)) consists in deducing a target hypothesis, from the evidence plus some general auxiliary hypotheses. Newton employed DI in his *System of the World* to deduce the inverse square law of gravitational attraction. The more general auxiliary assumptions that he used were: (a) Kepler's third law: $T^2 \propto r^3$, where T a planet's orbital period, and r the radius of its orbit (assumed circular); (b) Newton's laws of motion and, more specifically, their

consequence that the centripetal acceleration, a_c , of a planet moving at a circular orbit of radius r with tangential speed u is $a_c = \frac{u^2}{r}$; (c) the simple relation for circular motion $u = \frac{2\pi r}{T}$. From these assumptions, one obtains:

$$\begin{aligned} a_c &= \frac{u^2}{r} && \text{(from (b))} \\ &= (2\pi)^2 \left(\frac{r}{T}\right)^2 \cdot \frac{1}{r} && \text{(from (c))} \\ &= (2\pi)^2 \cdot \frac{r^3}{T^2} \cdot \frac{1}{r^2} \end{aligned}$$

that is:

$$a_c \propto \frac{1}{r^2} \quad (\text{since } \frac{r^3}{T^2} \text{ constant, from (a), and } (2\pi)^2 \text{ constant too})$$

There is a clear sense, then, in which (a special form of) the inverse square law of gravitation (i.e., the target hypothesis) can be deduced from the phenomena of planetary orbits (Norton 2000). And it is quickly seen that the premises of the argument in this case are more general and secure than the derived statement.¹⁷

Let us now explain how DI can be seen as supporting equation (2) as the better confirmed computational hypothesis for the LGMD neuron.

The form of $f(t)$ in (2) can be derived by the following, more general, auxiliary assumptions and experimental observations (Gabbiani et al. 1999, 1133):

- (a) $f(t)$ should depend only on information received at the retina; that is, θ and $\dot{\theta}$ (i.e., ψ).¹⁸
- (b) The firing rate at any moment t depends on the value of θ and $\dot{\theta}$ at the previous time $t - \delta$. That is: $f(t) = f(\theta(t - \delta), a\psi(t - \delta))$. A delay between stimulus and firing is theoretically expected owing to lags introduced by synaptic and cellular elements along the neuronal pathways converging onto the LGMD. In this case, the delay is *also* experimentally observed as a time difference between the moments that θ_{thr} is reached and the firing rate peaks. The parameters a and δ can be determined by means of linear regression of the experimentally obtained data points when measuring the dependence of the peak time, t_{peak} , on $\frac{l}{v}$; that is:

$$|t_{\text{peak}}| = a \frac{l}{|v|} - \delta \tag{5}$$

Thus, the parameters a and δ are identified as respectively the slope and intercept of the observed linear relation (Gabbiani et al. 1999, 1126).

¹⁷ That was a brief example of DI, owing to limitations of space in this paper, but we refer the reader who wants to better understand the logical structure of the method to Dorling (1973).

¹⁸ Technically, this possibility also includes the image edge acceleration, $\ddot{\theta}$. But this option was already excluded by the researchers based on precedent controlled experiments.

- (c) $f(t)$ should be of such a form that the experimentally observed linear dependence (5), between t_{peak} and $\frac{I}{v}$, should be satisfied for a variety of different θ_{thr} and δ values across different animals.

Based on these assumptions, Gabbiani et al. (1999) are able to mathematically derive Eq. (2) and show that it is the only functional combination of ψ and θ satisfying the desired properties. We will not present the mathematical steps here.¹⁹ What is important to emphasize for our purposes is that the above assumptions suffice to lead to a (computationally-relevant) hypothesis, which not only is not underdetermined, in any of the strong senses authorized by UTE, but, on the contrary, is mathematically determined. Thus, the determined hypothesis is as secure as the reports of the phenomena in assumptions (a)–(c), which now bear the inductive risk of the argument.

Implementation level

At this level, several hypotheses were considered as well. As we have said, from the three implementation-related hypotheses we mentioned in Sect. [What computational hypotheses are in play](#), the researchers put forward the one according to which the multiplication of the two terms in the kinematic part, $\psi(t - \delta) \cdot e^{-a\theta(t - \delta)}$, is implemented postsynaptically by means of dendritic summation of the excitatory and inhibitory inputs (the former logarithmically compressed) and followed by exponentiation (see, Eq. (4)). Grounds for adopting this log-exp hypothesis are a series of experiments and statistical significance tests (Gabbiani et al. 2002) as well as employment of a compartmental simulation model of the LGMD (Jones and Gabbiani 2010, 2012).

More specifically, Gabbiani et al. (2002) used a series of controlled experiments to disconfirm the presynaptic-inhibition hypothesis —the third of the three hypotheses mentioned in Sect. [What computational hypotheses are in play](#)— and then to directly test the log-exp hypothesis, by examining whether the exponentiation of $\log \theta - a\theta$ could occur while the membrane potential is converted into spike output by active membrane conductances. That is, by recording, in different animals, the connection between firing frequency and the intracellular membrane potential close to the spike initiation zone (SIZ).

Having the detailed results and methodological approaches of our case study under our belt, let us now return to addressing the main questions of this paper.

¹⁹ See the Appendix 3 of Gabbiani et al. (1999).

Determining the indeterminate? The confirmation methods against IC

Given that physical systems as “simple” —in terms of their input/output profile— as system \mathcal{P} (Table 3) can be computationally indeterminate, on what grounds are cognitive scientists justified in determining the computational identities of far more complex systems (such as neurons), given that the latter would be expected to be at least equally prone to IC?²⁰ To start articulating our proposed answer, we need to consider the question: could the LGMD neuron be seen as simultaneously realizing some other computation(s) as well, besides the one expressed by Eq. (2)? In the interests of brevity, we consider only the case of functional IC in this section. A brief discussion on the possibility of interpretative IC is provided in the Appendix.

Overcoming functional IC

Could the LGMD system be amenable to different groupings of its physical states, so that functional IC becomes a relevant concern? Answering this question is far from straightforward; it requires attending to the several physical interactions that take place concurrently in the system in order to see whether reasonable variant groupings of them can give rise to multiple formal/functional types of organization. We argue, however, that without using the confirmation methods identified by the UTE literature, there is no inherent reason to rule out such alternative types, while we show how the application of these methods to contextual clues guides the identification of the preferred formal/functional profile (resulting in the DI-derivation of Eq. 2).

To begin, the determined computational hypothesis about the LGMD posits a collective computation distributed over several thousand synapses. Furthermore, it comprises various particular mathematical transformations along the pathway from photoreceptors through the lamina and medullary neurons to the LGMD’s dendrites and its SIZ. For example, retinotopic inputs from individual facets in the eye become synchronized excitatory synaptic inputs to the LGMD through a mechanism that effects temporal coherence of the signals across individual facets. However, the synchronized excitatory input to the LGMD is not linearly related to the stimulus angular speed but follows a power-law transformation, which is also carried out at some point presynaptically along the pathway from the retina (Jones and Gabbiani 2012, 4930; see Fig. 7 top left). And yet, the effected change in the dendritic membrane potential, V_m , due to the excitatory inputs, rises much more gradually over time, owing to a further transformation that is well described by a logarithmic

²⁰ This is a claim about an intuition of course, thus open to objections. It is hard to provide a mathematical result that any system more complex than \mathcal{P} has to be indeterminate as well. But a recent result for a restricted domain indicates something towards this direction. Considering interpretative IC for Boolean functions, Fresco et al. (2021) show that IC becomes increasingly common as the system concerned computes functions with ever larger numbers of inputs. Specifically, the probability that a Boolean function with n inputs is *not* indeterminate is: $p_n = \frac{2^{2^{n-1}}}{2^{2^n}}$.

function (Fig.5). Since V_m is followed closely by the SIZ potential, this indicates that the logarithmic compression occurs locally within the dendrites (*ibid.*). At the same time, the inhibitory inputs to the LGMD follow the angular size in accordance with an approximate square-law transformation. As a result, the effect of the inhibitory input on the membrane potential shows a sigmoidal dependence on angular size (*ibid.*). Finally, the sum of the two signals (membrane potentials) becomes again transformed by the spike generation mechanism near the axon, in accordance to a power law close to an exponential (as in Fig.6). Figure 7 gives a schematic picture of most of these particular transformations.

The upshot of the above is that a variety of mathematical operations are implemented locally at different parts of this large visual system (including the pathway from the retina to the LGMD-DCMD neurons). Let us assume momentarily that we have found out the above biophysical interactions but we have not yet figured out what the LGMD's functional organization and computed function are. By focusing on different local interactions, and by grouping them together under different functional descriptions, one can be led to multiple formal/functional organizations (even at different scales). For example, by looking at the synchronized excitatory synaptic input to the LGMD in isolation, one sees a change in the excitatory input that is not linearly correlated with increase of angular speed (top left in Fig.7; the relation approximates a square- or third-power law), thereby not directly supporting a computation involving information about the angular speed of the looming object, since the latter quantity is never directly represented within the neural populations presynaptic to the LGMD. Based on this group of interactions, a determined formal/functional structure might just posit some computation related to a power-law transformation. Nevertheless, the said non-linear correlation becomes immaterial when the point of view subsumes the logarithmic compression carried out postsynaptically (figs. 5 and 7 middle left), since the logarithm of speed raised to any power is simply equal to the logarithm of speed multiplied by that power's exponent (i.e.: $\log \psi^a = a \log \psi$); thereby counteracting the non-linearity between angular speed and excitatory input (Jones and Gabbiani 2012). A variant grouping of the interactions then, which also takes into account the logarithmic transformation, changes the possibilities of formal/functional structures, since the fact that the angular velocity is never directly represented presynaptically becomes inessential (the foregoing non-linearity becomes canceled out).

Furthermore, an alternative, extended, grouping that incorporates the spike generation mechanism near the axon changes again the picture, since the (exponential) transformation of the net membrane potential into firing rate that occurs in the SIZ can now give rise to the computational hypothesis captured by Eq. (2). This is, of course, by virtue of the equivalence (4), which now allows of abstracting away from the logarithmic and exponential transformations (figs.5 and 6) at the computational and algorithmic levels of description. Thus, the dynamical behavior observed, e.g., in Fig.5 can be given, under this variant grouping, yet again a different computational interpretation, since the computational description (2) can only arise insofar as the lower-lever dynamical characteristics —depicted in figures 6 and 5— are typed together under this different computational taxonomy.

Despite the existence of several possibilities about how to carve up the range of transformations occurring in our system, the researchers ultimately single out a

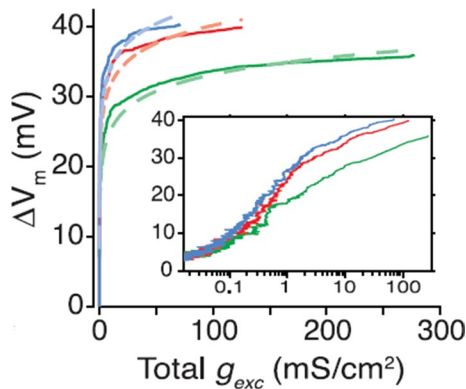
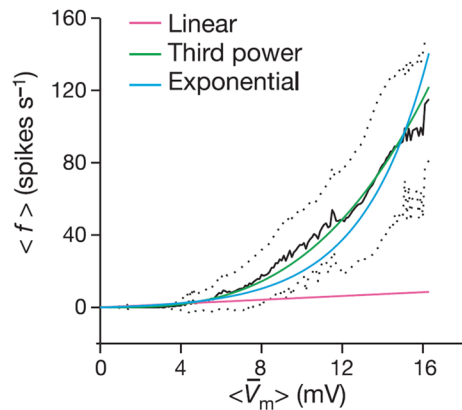


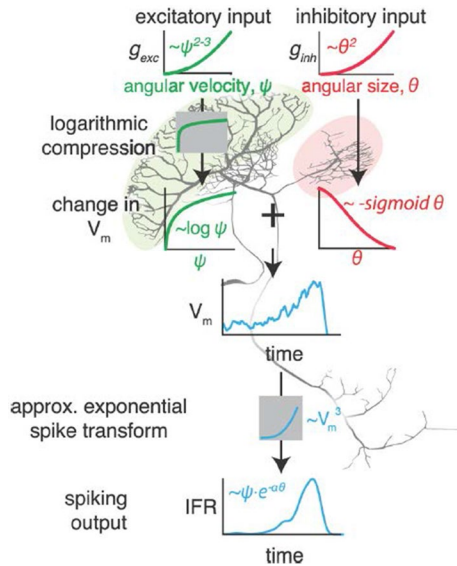
Fig. 5 The relationship between $g_{exc}(t)$ and the resulting membrane potential near the spike initiation zone (SIZ). Colors show responses to looming stimuli with different $\frac{1}{v_{rel}}$ values (green: 10ms; red: 40ms; blue: 80ms). Solid lines show the observed relationship, while dashed, lighter lines show fitted functions of the form $V_m = a \log(bg_{exc})$. The inset shows the data on a logarithmic x -axis, so that the relationship becomes approximately linear. (From Jones and Gabbiani (2012), with permission.)

Fig. 6 Fit of mean instantaneous firing rate, $\langle f \rangle$, as a function of mean, median filtered membrane potential (mean \pm s.d.; solid and dotted black lines) with linear, third-power and exponential models. (From Gabbiani et al. (2002), with permission.)



unique one as better characterizing the system computationally. The identified computation relates the time course of a *compound* conductance —distributed over the entire dendritic tree of the neuron— with the membrane potential at the SIZ. But what really makes this particular partitioning justified? Why not conclude (say) that the LGMD dendrites compute a logarithm, and its SIZ implements an exponential function, but rather that the whole neuron computes a plain multiplication of two terms? Or, why assume, in the first place, that the LGMD computes a continuous function and not some Boolean operation, in direct analogy with system \mathcal{Q} (in

Fig. 7 Schematic of the computation performed by the LGMD.
 (From Jones and Gabbiani (2012), with permission.)



Sect. [The IC in more detail](#)) which can be seen as implementing either a multiplication or an AND/OR logical function under different circumstances?²¹

A possible reply to the first of these questions might be that since a neuron is an autonomous unit, it is natural to be computationally treated as an autonomous unit too (the ‘point-neuron’ as is sometimes called in the relevant literature). However, such a response would not be satisfactory. There are cases of multi-neuron systems that can be seen as collectively implementing some specific formal/logical structure. It also seems likely that there are cases of computations performed by only specific parts of neurons.²² Thus, an exclusively bottom-up (i.e., context-independent) investigation does not suffice to decide between these options. It is the additional, non-trivial clues provided by the context, as well as the application of the confirmational techniques from Sect. [Responses to UTE](#) to these clues, that support the hypothesis that the logarithmic dendritic compression and the exponential SIZ transformation should be grouped together in our case. Furthermore, as the example of system *Q* indicates, there are systems that can be seen as simultaneously implementing continuous (e.g., a multiplication) as well as Boolean functions. How do we exclude that the LGMD is not such a system or, if it is, that a Boolean gate is not the

²¹ Strictly speaking, the two questions reflect different forms of functional IC, since one has to do with changing the boundaries of the considered computing system, while the other concerns different groupings of interactions within the same system. But, although this distinction is meaningful from an ontic point of view, it practically vanishes in an epistemic context, since the exact boundaries of the computing unit are typically not known in advance, and identifying them is also part of the process of determining the computational profile of the studied neuronal structure. The proposal of this paper (confirmational devices plus contextual clues) addresses both issues as well.

²² See, e.g., Poirazi et al. 2003; London and Häusser 2005; Gidon et al. 2020.

appropriate computational profile to single out but it is Eq. (2) instead? Another possible reply might be that such a complicated neuronal system as the LGMD would not be reasonably seen as just implementing a single Boolean gate; such a hypothesis might seem too simplistic to even get off the ground. But Boolean functions had been hypothesized as the main computed operations by neurons in the past, and they have reappeared as likely computations of complex systems, in recent research.²³ These considerations show that there is no principled reason to exclude alternative groupings of the many concurrent physical transformations of signals that may give rise either to partial computational hypotheses at different parts of the LGMD (e.g., a logarithmic computation at the dendrites) or to additionally implemented (but not actually exploited) Boolean operations by virtue of coarser-grained groupings of the input-output relations.²⁴

To make the connection with the earlier examples of IC clearer, recall the example of system \mathcal{Q} , which can be seen as either implementing AND/OR or multiplication of its two inputs. The identified complex dynamical behavior of the LGMD, consisting of the several transformations occurring at its different parts and depicted in Fig. 7, can be seen as analogous to the dynamical behavior of \mathcal{Q} , depicted in Table 6. Accordingly, the determined computation of a product, captured by Eq. (2), is analogous to the determined computation of a product, captured by Eq. (1). In both cases, the determined computation results from a specific way of grouping together the several input/output relations (which were themselves identified by means of measurements, controlled experiments, and confirmation methods), as a certain formal/functional profile. And in both cases, the (additional) application of the confirmation stratagems to the contextual clues guides the determination of the relevant hypotheses in such a robust way that any alternative groupings might seem from the beginning highly implausible or irrelevant (so that the system might appear from the outset as not suffering from indeterminacy). But this should not mislead us in believing that the system has indeed a unique computational profile. What it rather shows is that the confirmational methods achieve two goals in one fell swoop: determine dynamic behaviors *and* exclude the non-preferable functional/formal profiles in the given context.

Let us see how this works specifically for our case study. Putting the computational investigation of the LGMD-DCMD system in the context of vision provided the following crucial contextual clues: (a) The LGMD-DCMD system is linked *as a whole* with escape behaviors based on visual stimuli (based on previous research). (b) The identified computation should be such that its form depends only on information received at the retina. (c) Since the inputs stem from stimuli at the retina, which is a far distance from the SIZ, the firing rate at any moment t should depend

²³ See, for example, (Gidon et al. (2020, 86), where the research team recently identified concurrent Boolean operations at different parts of layer 2 and 3 pyramidal neurons of the human cerebral cortex. Specifically, the team put forward the hypothesis that the dendrites of these cells compute XOR operations while the soma and the tuft and basal dendrites compute AND/OR operations.

²⁴ Getting bogged down with the actual measurements of the values involved in the LGMD study in order to construct alternative coarser-grained groupings for a Boolean profile would unnecessarily complicate this paper. It suffices for our purposes that such a carving up is in principle possible, as the system \mathcal{Q} shows.

on the value of θ and $\dot{\theta}$ at an earlier time $t - \delta$. These clues were crucial in identifying the functional taxonomy of the system, and all the more so given that the last two assumptions —(b) and (c)— were also the first and second assumptions in the DI-derivation of Eq. 2 (Sect. [How computational hypotheses are confirmed](#)).

Before closing this section, we emphasize that it is not only against functional IC that the context makes a crucial difference. Contextual clues play an implicit yet decisive role in overcoming interpretative IC too. For this case, we refer the interested reader to the brief discussion in the Appendix, where it can be seen that the clues obtained by the assumption of a computation within a visual context underpin the arguments against interpretative IC as well.

Conclusions

We examined a case study from computational neuroscience where computational hypotheses concerning a large neuronal system that involves several mathematical transformations were singled out. We can distill the process into a scheme along the following lines.

Suppose that we aim to model, in computational terms, a physical system of unspecified complexity; be it either a simple voltage-in/voltage-out system (as the system \mathcal{P} in Fig. 1) or even a large (set of) neuron(s). The physical interactions that constitute the dynamical profile of the system are determined by means of the various methods discussed in Sect. [Responses to UTE](#). So far this is an effective approach against UTE concerns, but not yet against IC (i.e., it is effective to characterize the system physically, but not yet computationally). Functional and interpretative IC are grounded, respectively, in the existence of different possibilities of grouping the (well-determined) physical interactions and of interpreting any (well-determined) formal/physical states. But not all hope is lost. Our good old friends, the conceptual devices for confirming empirical hypotheses can provide reasonable constraints on what the *appropriate* physical groupings and/or interpretations within a particular context are. As the LGMD study suggests, such constraints can be rigorously determined by (a) obtaining evidential clues from the context in which the computational system operates, and, subsequently, (b) employing the various stratagems of the kind that make scientific theorizing and hypothesis testing generally possible, subsuming also the external, contextually-obtained evidence.

Finally, it is useful to clarify where this proposal stands with respect to the ongoing debate about computational individuation (Sect. [The current debate on computational-individuation and the relation between IC and UTE](#)). It might seem that our thesis is a reiteration of what we called the ‘extrinsic’ view, which subsumes contextual clues —functional, semantic, or whatnot— under the computational identity itself. But, as we emphasized in Sect. [The current debate on computational-individuation and the relation between IC and UTE](#), our actual concerns are with the epistemic, rather than the ontic, aspect of computational individuation. That is, we are concerned with its heuristic methodology (how we find out the computational profile in practice) and its justification (why it is warranted to select the specific profile from all the simultaneous

alternatives). In other words, we are merely interested in the scientific discovery of computational hypotheses (profiles) in the actual practice. And scientists do rely in practice on contextual clues when coming up with computational hypotheses about cognitive systems. So although we do not take an explicit stance here on the type of the crucial contextual clues —functional, semantic, or others— for computational individuation, one can still argue that our proposal weighs in on the ontic debate and tips the scales against the intrinsic view of individuation.

Nevertheless, we do not take the extra step of making such a claim here. For it is possible that the supporter of the intrinsic view could, perhaps, make a case that the question of how scientists discover and justify computational profiles is a separate one from the question of what the (metaphysical) nature of computational individuation is. Thus, we remain neutral about the ontic debate and propound only an *epistemic* claim. Regardless of whether the computational profile itself is ultimately context-dependent or not, the conceptual and mathematical toolbox for hypothesis confirmation provides also a sufficient underlying framework for determining computational profiles in the presence of functional and/or interpretative computational indeterminacy —insofar as it is sufficient for determining explanatory hypotheses against any UTE concerns.

Appendix: the confirmation methods against interpretative IC

Since the determined formal/functional profile of the LGMD system does not eventually involve Boolean functions, the relevant scenario of potential interpretative IC would concern an indeterminacy about the interpretation of certain physical quantities and states. Recall from Sect. [What computational hypotheses are in play](#) that via a series of controlled experiments, aimed at the implementation level, Gabbiani et al. (2002) determined that the inhibitory input (Fig. 2) encodes information about size (angle subtended in the retina by the moving object), while the excitatory input encodes information about motion (angular edge velocity). But how are these interpretative hypotheses justified, given that there is nothing in the physics of the system itself that concerns angles, velocities, and whatnot, besides voltages, excitations, inhibitions, spikes, etc.? It is, after all, certainly conceivable that the same physical quantities —considered without the (empirically acquired) contextual knowledge that the LGMD neuron participates in a visual task— could have been interpreted as encoding different parameters of the external world, such as (say) the instantaneous phase and frequency of some sinusoidal signal. Since the underlying formal/functional structure (i.e., Eq. (2)) would remain the same,²⁵ such a possibility of multiple interpretations would automatically signify the possibility of interpretative IC.²⁶

²⁵ The instantaneous phase of a sinusoidal signal is the time derivative of its frequency; thus the formal relations are the same between phase–frequency and angle–angular edge velocity.

²⁶ There is disagreement about whether such a change of computational content (e.g., from visual to auditory tasks) would actually affect the system’s computational identity. The issue has been discussed by Burge (1986), Egan (1995), and Shagrir (2001), and it ultimately turns on one’s precise views of ‘computational implementation’ and ‘individuation’. This, however, does not affect our basic point here.

So let us see how the confirmation devices help to justify that the interpretative hypothesis about encoded size and motion is the preferable one. We need to consider the formal properties of Eq. (2). Since g is time-independent, the time course of the LGMD firing rate depends entirely on the argument of g ; that is, the kinematic part $z(t) = \psi(t - \delta) \cdot e^{-a\theta(t-\delta)}$. An increase in ψ (recall that $\psi = \frac{\theta}{2}$) causes $z(t)$ to increase too, whereas an increase in θ causes it to decrease (owing to the minus sign in the exponent). At the onset of the approach, the ψ term—angular edge velocity—has a dominant effect in the increase of the firing rate; as the object is still far away, it subtends a small angle to the retina, making the contribution of the $e^{-a\theta(t-\delta)}$ term small. As time goes by, the contribution of the latter part gains dramatic significance (due to its exponential dependence on θ), thereby leading to a rapid decrease of the firing rate eventually (Fig. 3). As a result, the mechanics of Eq. (2) indicates that it would be natural to assume that the behavior of ψ and θ gets encoded in the neuron by some physical quantities that have competing contributions to the net firing rate of the neuron. Such quantities are excitation and inhibition; thus, the plausible interpretative hypothesis adopted by Gabbiani et al. (1999, 1133) is that ψ acts as an excitatory term, whereas θ acts as an inhibitory term. Additionally, this hypothesis seems suitable to explain a long-known apparent excitatory and inhibitory effect of motion and size on the LGMD response (e.g., Palka 1967).

To conclude, here is the argument about the confirmation methods to overcome interpretative IC. The determined formal/functional organization of the system (i.e., Eq. (2)) admits of more than one interpretation of its abstract/functional states (values of ψ and θ). The confirmation devices make it possible to single out and justify the interpretative hypothesis that the inhibitory and excitatory inputs encode, respectively, size- and motion-dependent information. The warrant for this hypothesis is provided by converging evidence from diverse domains; that is, existence of evidential agreement that would seem unlikely, were this hypothesis false (cf., Sect. Responses to UTE). This is indeed the case in our study, because none of the three assumptions (a–c) that are used in deriving Eq. (2) states or presupposes anything about the excitatory and inhibitory effects of motion and size on the LGMD response. In other words, the adoption of our interpretative hypothesis is based on evidential support that stands on two independent lines of reasoning: First, Eq. (2) is a very successful hypothesis—it fits well with the experimental profile of the LGMD response (Fig. 3)—which lends itself very naturally to interpreting its terms as encoding excitatory and inhibitory effects, and, concurrently, it is formally derived from assumptions that *say nothing* about such excitatory and inhibitory effects (no mapping between the mathematical terms and any excitatory or inhibitory inputs is assumed). Second, motion and size have long been observed to have an excitatory and inhibitory effect on the LGMD response.

Both of these, independently warranted, hypotheses, then, indirectly support the further hypothesis that the excitatory and inhibitory inputs encode, respectively, information about motion and size. And since both of the two independent hypotheses are warranted on grounds offered by the kind of methods we have discussed in the context of UTE, it is the same methods that lend confirmational support to this interpretative hypothesis as well.

Acknowledgments We thank Jack Copeland for useful discussions on the topic of this article. We are also thankful to several edanonymous referees for constructive comments and suggestions. Earlier versions of this article were presented at the HAPOC5 meeting in Bergamo and at the Indeterminacy and Underdetermination Workshop at the University College, Dublin. We would like to thank the audiences of both events.

Funding This research was supported by the Israel Science Foundation Grant (830/18).

Declarations

Conflict of interest The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

Consent for publication We certify that the submission is original work and is not under review at any other publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Burge T (1986) Individualism and psychology. *Philosophical Rev* 95(1):3–45
- Coelho Mollo D (2018) Functional individuation, mechanistic implementation: the proper way of seeing the mechanistic view of concrete computation. *Synthese* 195(8):3477–3497
- Crupi V (2016) Confirmation. In: Zalta EN (ed) *The stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/spr2021/entries/confirmation/>
- Dewhurst J (2018) Individuation without representation *Br J Philosophy Sci* 69(1):103–116. <https://doi.org/10.1093/bjps/axw018>
- DiSalle R, Harper W, Valluri S (1994) General Relativity and empirical success In *Proceedings of the Seventh Marcel Grossman Meeting on recent developments in theoretical and experimental general relativity, gravitation, and relativistic field theories*, pp 470–471
- Dorling J (1973) Demonstrative induction: its significant role in the history of physics. *Philosophy Sci* 40(3):360–372
- Dorling J (1990) Reasoning from phenomena: Lessons from Newton In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1990*:197–208
- Dorling J (1995) Einstein's methodology of discovery was Newtonian deduction from the phenomena, In *The Creation of Ideas in Physics: studies for a methodology of theory construction*, ed. Leplin J, 97–111 Springer https://doi.org/10.1007/978-94-011-0037-3_6
- Dretske FI (1988) *Explaining behavior: reasons in a World of Causes*. MIT Press, Cambridge
- Egan F (1995) Computation and content. *Philosophical Rev* 104(2):181–203
- Egan F (2012) Metaphysics and computational cognitive science: let's not let the tail wag the dog. *J Cognit Sci* 13(1):39–49
- Fresco N (2021) Long-arm functional individuation of computation. *Synthese* 199:13993–14016. <https://doi.org/10.1007/s11229-021-03407-x>
- Fresco N, Copeland BJ, Wolf MJ (2021) The indeterminacy of computation. *Synthese* 199:12753–12775. <https://doi.org/10.1007/s11229-021-03352-9>
- Fresco N, Miłkowski M (2021) Mechanistic computational individuation without biting the bullet. *Br J Philos Sci* 72(2):431–438

- Gabbiani F, Krapp HG, Koch C, Laurent G (2002) Multiplicative computation in a visual neuron sensitive to looming. *Nature* 420(6913):320–324
- Gabbiani F, Krapp HG, Laurent G (1999) Computation of object approach by a wide-field, motion-sensitive neuron. *J Neurosci* 19(3):1122–1141
- Gidon A, Zolnik TA, Fidzinski P, Bolduan F, Papoutsi A, Poirazi P, Holtkamp M, Vida I, Larkum ME (2020) Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* 367(6473):83–87. <https://doi.org/10.1126/science.aax6239>.
- Harbecke J, Shagrir O (2019) The role of the environment in computational explanations. *Eur J Philosophy Sci.* <https://doi.org/10.1007/s13194-019-0263-7>
- Harper W (1990) Newton's classic deductions from phenomena PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1990(2):183–196. <https://doi.org/10.1086/psaprocbie.nmeetp.1990.2.193067>
- Harper W (1997) Isaac Newton on empirical success and scientific method. In: Earman J, Norton JD (eds) *The cosmos of science: essays of exploration*. University of Pittsburgh Press, pp 55–86
- Hatsopoulos N, Gabbiani F, Laurent G (1995) Elementary computation of object approach by a wide-field visual neuron. *Science* 270(5238):1000–1003
- Jones PW, Gabbiani F (2010) Synchronized neural input shapes stimulus selectivity in a collision-detecting neuron. *Curr Biol* 20(22):2052–2057. <https://doi.org/10.1016/j.cub.2010.10.025>
- Jones PW, Gabbiani F (2012) Logarithmic compression of sensory signals within the dendritic tree of a collision-sensitive neuron. *J Neurosci* 32(14):4923–4934
- Laudan L (1990) Demystifying underdetermination. In: Savage CW (ed) *Scientific theories, minnesota studies in the philosophy of science*, vol 14. pp 267–297
- Laudan L, Leplin J (1991) Empirical equivalence and underdetermination. *J Philos* 88(9):449–472
- London M, Häusser M (2005) Dendritic computation. *Annu Rev Neurosci* 28(1):503–532. <https://doi.org/10.1146/annurev.neuro.28.061604.135703>
- Maley CJ, Piccinini G (2017) A unified mechanistic account of teleological functions for psychology and neuroscience. In: Kaplan DM (ed) *Explanation and Integration in Mind and Brain Science*. Oxford University Press, Oxford
- Marr D (1982) *Vision: a computational investigation into the human representation and processing of visual information* NY, WH, Freeman and Company
- Matthews RJ, Dresner E (2017) Measurement and computational skepticism. *Noûs* 51(4):832–854
- Millikan RG (1984) *Language, thought, and other biological categories: new foundations for realism*. MIT Press, Cambridge
- Norton JD (1995) Eliminative induction as a method of discovery: how Einstein discovered general relativity. In: Leplin J (ed) *The creation of ideas in physics studies for a methodology of theory construction*. Springer, New York. https://doi.org/10.1007/978-94-011-0037-3_3
- Norton JD (2000) How we know about electrons. In: Nola R, Sankey H (eds) *After Popper, Kuhn and Feyerabend: recent issues in theories of scientific method*. Springer, Dordrecht, pp 67–97
- Norton JD (2008) Must evidence underdetermine theory? In: Carrier M, Howard D, Kourany JA (eds) *The challenge of the social and the pressure of practice: science and values revisited*. University of Pittsburgh Press, Pittsburgh, pp 17–44
- Palka J (1967) An inhibitory process influencing visual responses in a fibre of the ventral nerve cord of locusts. *J Insect Physiol* 13(2):235–248. [https://doi.org/10.1016/0022-1910\(67\)90151-5](https://doi.org/10.1016/0022-1910(67)90151-5)
- Papayannopoulos P, Fresco N, Shagrir O (2022) On Two different kinds of computational indeterminacy *Monist* 105(2):229–246. <https://doi.org/10.1093/monist/onab033>
- Piccinini G (2015) *Physical computation: a mechanistic account*. Oxford University Press, USA
- Poirazi P, Brannon T, Mel BW (2003) Pyramidal neuron as two-layer neural network. *Neuron* 37(6):989–999
- Psillos S (1999) *Scientific realism: how science tracks truth*. Routledge, UK
- Rescorla M (2014) A theory of computational implementation. *Synthese* 191:1277–1307. <https://doi.org/10.1007/s11229-013-0324-y>
- Salmon MH et al (1992) *Introduction to the philosophy of science*. Prentice-Hall, Hoboken
- Shagrir O (2001) Content, computation and externalism. *Mind* 110(438):369–400
- Shagrir O (2012) Computation, implementation, cognition. *Minds Mach.* 22(2):137–148. <https://doi.org/10.1007/s11023-012-9280-4>
- Shagrir O (2020) In defense of the semantic view of computation. *Synthese* 197:4083–4108

Sprevak M (2010) Computation, individuation, and the received view on representation. *Studies in history and philosophy of science part A* 41(3):260–270. <https://doi.org/10.1016/j.shpsa.2010.07.008>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.