The Role of Analog Computation in Path Integrating Behaviour of the Desert Ant

A Defence of the Explanatory Credentials of Connectionism in Cognitive Science

Robert Peter Farquharson

B. Psych. Sci., B.A. Hons. (Philosophy)

Department of Philosophy, School of Humanities

The University of Adelaide

Thesis submitted in fulfilment of the requirements for the degree of Master of Philosophy

December | 2018

Abstract

Declaration

Acknowledgements

Chapter 1

Introduction

Chapter 2

Computation & the Philosophical Foundations of Cognitive Science

Chapter 3

Path Integration & Polarisation-Sensitive Vision in the Desert Ant

Chapter 4

The Explanatory Demands of Desert Ant Navigation

Chapter 5

Defending the Representational & Computational Credentials of Connectionism

Chapter 6

Conclusion

Abstract

What is the relationship between the mind and the brain? Cognitive science is a unique discipline of inquiry that explores this question. Its uniqueness results from a commitment to investigating and understanding cognitive systems as information processing systems. Traditionally, the commitment to information processing has entailed two further positions regarding the nature of "information", and how exactly it is "processed" in a medium like the brain: representationalism, and computationalism. Taken jointly, these commitments have equipped cognitive science to understand the biological mechanisms by which organisms identify and adapt to variations in their environment. That is, cognitive science is in the business of naturalistically explaining intelligence as the leveraging of internal representations in computational operations towards some goal.

Much debate has surrounded the exact manner in which notions like representation and computation are to be cashed out. This debate has mostly centred around two competing accounts: classicism and connectionism. The present thesis is motivated by a desire to fully separate these two competing frameworks, because there is much confusion about whether this is possible. I take an affirmative position on that issue: connectionism is indeed an independent and unique computational framework.

Both classicism and connectionism are representational and computational accounts. What separates them is the kind of representations they appeal to, and thus the species of computation they perform over those representations. Connectionism appeals to structural representations, defined over the intrinsic physical properties of the representing medium. Connectionism is thus an analog computational framework: structural representations sustain physical analogies with their task domain.

The now ubiquitous 'classical' framework, on the other hand, operates over symbolic representations, and is thus a digital computational framework. Much ink has been spilled in assessing the theoretical differences and advantages/disadvantages of the two rival accounts. The present work is, instead, a more hands-on comparative analysis — using an illustrative case study to demonstrate the practical differences in the application of these accounts.

A contemporary and influential cognitive scientist of the classical stripe, C. R. Gallistel, has criticised connectionism by (putatively) showing how it

can't explain the rudimentary cognitive phenomenon of Desert Ant path integration and dead reckoning behaviour. I take up this challenge, by 1) showing how Gallistel has misrepresented connectionism as a symbolic computational framework, and then 2) demonstrating exactly how a "true" connectionist account can interpret and model the neurobiology involved in path integration.

Using the case study of Desert Ant behaviour, the true appeal and independence of the connectionist framework can be seen, separating it from its classical, symbolic rival. Abandoning symbolic computation, in favour of a structural and analog approach, secondarily shines new light on many theoretical issues that are nested within the modelling and understanding of animal navigation tasks. These have further implications for how we think of animal behaviour, computation, and intelligence more generally.

Structural representations and analog computation can handle the explanatory load of the rudimentary example of animal intelligence in the Desert Ant. They also offer up robust and promising lines of inquiry for a naturalistic cognitive science to pursue.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Robert P. Farquharson

Signature:

Date:

Acknowledgements

Chapter 1
Introduction

A Philosophical Survey of Minds in a Material World

"[If] it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying… if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world."

- *Jerry Fodor*

1.1 Between Democritus & Descartes

All good philosophy begins with two opposing intuitions. Both seem equally convincing and true, and yet appear to be mutually exclusive. Consider the following:

1) The universe is exhaustively composed of physical matter, e.g. atoms and their purposeless causal interactions.

2) Mental states have an ostensibly non-physical character, e.g. they preside over purpose-driven causal interactions — in accordance with their individual contents.

Call the first intuition the *Naturalist Constraint*. The Naturalist Constraint is an expression of current scientific knowledge, its success, and its undeniable progress. It encompasses both an ontological picture of the universe, and an epistemic commitment regarding how we can fill in that picture. We're committed to poking and prodding the universe in a certain way whenever we build upon our old theories, or when we launch new investigative pathways. A definitive encyclopaedia of the objective universe, according to naturalism, will include only those things made of physical parts. This includes things which may result from a specific organisation or interaction of such physical parts, but at bottom can be said to supervene on their physical parts. Anything not discoverable by systematic physical interaction, and thereby not describable using the language of the scientific point of view, will be eliminated. Colloquially, naturalism is often defined in contrast with its opposite: no spooky stuff — nothing *super*natural.

Call the second intuition the *Cognitivist Constraint*. The Cognitivist Constraint is an expression of the intimate and undeniable knowledge we all have of our own subjective realities. We experience the world as being a certain way, and interact with it in ways that, at least to us, are *reasonable* given the way we perceive it as being. Mental states act as the nexus between us and the world precisely because they are "*intentional*" in this way: they are *about* certain things. Any discrete mental state, and its causal powers, can therefore be individuated by its *contents* — the particular way that some thing is rendered as being. Everyday life is littered with evidence of this reality. Our beliefs, desires, and perceptions all causally mediate our behaviours in different ways. They're an air gap between input from the world and the output we give back, but notice that they perform this function specifically with respect to what they are individually about.

The tension between 1) and 2) has been famously difficult to resolve. Naturalism explicitly endorses a thorough-going *materialism*, but there are no molecules of meaning, no atomic units of aboutness to straightforwardly import from the material sciences to explain our intuitions of *intentional realism* — that the contents of mental states (beliefs, desires, attitudes, etc.) are real and drive reasonable behaviour. Two extreme responses to this problem are to jettison one or the other constraint.

Some generic brand of *eliminativism* cautiously falls on the side of the naturalist. It suggests that whatever "mental states" may end up being, they're not what we think they are. At best, current talk of purposive action driven by contentful mental states is a useful heuristic, but these categories should not enjoy a real place in our scientific image of the world. All of the causal and explanatory work will ultimately be described by sole reference to the material substrate of our brains, making mentalistic

ascriptions irrelevant. Democritus is perhaps the earliest exponent of an ardent and eliminative naturalism, captured by his pithy claim that there are only "atoms and the void". On the other hand, some generic brand of *dualism* rejects the Naturalist Constraint, invoking a non-physical, mental substance to explain the reality of mental states. Science may not have the appropriate language for explaining mentality, but that's because we need an entirely new language. Descartes is the most influential progenitor of this kind of radical "substance dualism", where the mind, and its mentalistic properties, is in every way separate from the body.

Both these extremes have been well critiqued over the years, one for leaving something critical out, the other for adding too much. Eliminating mentality, or at least deflating it to some illusory status, is unpalatably narrow because it runs counter to our subjective experience. Throwing away notions like 'meaning' or 'aboutness' wholesale would seem to leave us powerless to carve out a privileged place in the world for *intelligent* behaviours — those purposive actions performed by creatures such as ourselves, non-human animals, and perhaps to a lesser extent even plants. We undoubtably want to separate intelligent behaviour from the merely brute-causal world of inanimate objects, but we want a theory that explains how that is possible, without just hand-waving about increased scale or complexity; it must be fundamentally special in some way.

The dualist position is conversely unpalatable because it is too broad. We have no idea how a non-physical, mental substance could causally interact with the physical substance that makes up the rest of the universe. If mental states mediate our behaviour, and they do this via our brains, then it must still interact with the physical substrate of our bodies. Without some theory for explaining how this works, then the mental substance seems to be a difference that makes no difference at all. A bias for parsimony would dictate that we should simply leave it out of our world view.

As a matter of methodology, we should be aiming somewhere in the middle of these polar extremes. To rehabilitate the notions of meaning or intentionality, perhaps for now we can start by rescuing intelligence from the naturalist's razor. All mental states must be intentional; definitionally, they are *about* something. However, not all intentional states need be mental, i.e. conscious. Exonerating some core features of mental states from their guilt-by-association with *consciousness* can allow us to keep studies of intelligence within the naturalistic domain, without flatly denying the reality of something that seems plain for all to see; that we behave in certain ways because of the manner in which the world is rendered to us. We can demote the '*Hard Problem*' (see Chalmers, 1995) — why some intentional states are also conscious, *mental* states— to a lower position on our scientific to-do list, and concentrate on how it is that all mental states get to be intentional states. 'Cognitive' can then be synonymous with 'intelligent', without necessarily bringing the additional metaphysical baggage of 'conscious' along with it. The history of scientific progress, after all, has some promising precedents in carving out such middle paths.

The *élan vital* that was once thought to impel living organisms forward was eventually replaced by a purely naturalistic account of self-organisation, heredity, and complexity. As we understand more and more about the material substrates that make up the mechanisms of life, e.g. organic chemistry and biology, an extra supernatural category becomes irrelevant. There is indeed something special about living things that the old theory recognised, e.g. self-propagation, reproduction, familial resemblance,

and so on. The modern theory that superseded it, however, brought those special features into the fold of natural science without deflating their specialness. Indeed, I'm inclined to say they were made *more* special by a deeper understanding of their underlying mechanisms. Similar demystification projects have occurred across the board in the natural sciences. The invisible forces that govern the universe at its largest and smallest scales have all been elucidated by modern scientific frameworks like classical mechanics, quantum mechanics, cosmology, and so on.

In this tradition, somewhere between Democritus' "atoms and the void", and Descartes' non-physical substance, a modern cognitive science should endeavour to naturalise the study of intelligent action, and explain how meaning-rich mental states can guide it. However, it must do so without throwing the cognitive baby out with the supernatural bathwater.

1.2 Reasons & Representations

The task for a modern cognitive science is to fill the explanatory gap between Democritus and Descartes: to naturalise notions like *meaning*, *aboutness*, and *reasons*, without doing undue violence to their seemingly special character. To narrow the problem, we must first wade through some philosophical marshes. Just like the manner in which the naturalisation of life occurred, we should set up a theoretical landscape within which we can sharpen up what we think is so special and different about mental states and their causal influence on behaviour. With a clearer image of what to search for, we can then begin probing for potential mechanisms. Consider a typical morning in my head:

> What time/day is it? What time does my tram leave? What's my contingency plan if I miss my tram?

The relative success with which I navigate my morning's obligations will firstly depend on what I *believe* to be the answers to those thoughts. Secondly, it will depend on how well those beliefs match up with the external state of affairs around me. This is where the first flag needs to be planted, in recognising and emphasising that the primary manner in which the causal influence of mental states seems to be so distinctly special is that they aid in *solving problems*.

Depending on the day, I need to be at particular places at particular times. Accordingly, I have to link up with my transport at an appropriate time, given that I desire to meet those particular demands. This problem involves the use of mental states with distinct contents, that depict the world as being in particular ways. For example: the belief that it is Friday is different to the belief that it is Saturday; that a tram that leaves at 8:15am has different consequences to another that leaves at 8:45am; that arriving late has different consequences to arriving on time. All of these beliefs are distinct, discrete, and different to one another. The interactions among these distinct beliefs are what cause me to navigate the world in a distinct way, to bring about a particular state of affairs. I catch the 8:15am tram, and I arrive on time (mostly).

Importantly, if something goes wrong, I am equally able to adjust my behaviour in the face of the unexpected. If I wake up late, I may drive to uni instead. If I've been absent minded and turned up to class only to realise it's Saturday, I won't simply stand there and teach to an empty room. I'm able to take in new information and act accordingly. If I'm late, I *update* my beliefs about time and transport,

but *only* when I am late. When I'm in error about what day it is, I slink home lamenting that I could have slept in, but *only* when I'm in error. The way my mental states can change and operate together, internally affecting one another in my head, are ultimately what render my behaviour flexible and appropriate (or not) given the state of affairs in the world around me.

It has been standard in the philosophical literature to think of this kind of behaviour that I exhibit as special, and as intelligent, because it is reason-driven. I commit to particular actions for the reason that they will, by my lights, allow me to reach my ends. But reasons don't exist in the world, because reasons are just descriptions that map some states with other covariables. However, to act *because of*, or *for* a reason, just requires the capacity to possess some internal goal-state; a belief about the way the world is or should be. Reason-driven action involves engaging in a behaviour because you have a secondary belief that the behaviour will bring about that goal-state; you are privy to the mapping between the behaviour and its outcomes because the mapping is part of your subjective perspective of the world. If I take it to be the case that *A* brings about *B,* and I desire *B*, then I'll perform *A*.

This is a standard hallmark of intelligence: the selective response to certain aspects of the environment in the service of some some goal-state(s), i.e. problem solving, in its broadest construal. Approaching the study of intelligence through the lens of internal reason-driven action also pays enormous explanatory dividends. To paraphrase Fodor (1987: p. 6), to find out where I will be on any given morning simply *ask me*. One thereby discovers what my subjective world-view is like. One could probably give a highly detailed physical explanation of where my body will be at some future time by appealing to physical mechanics and so on, but not without considerable effort (if at all). In this way, "our best science of middle-sized objects" (*ibid.*) loses out, in terms of explanatory frugality and power, to the simple appeal to internal states and their contents. I can solve problems, and order my behaviour, in ways that more "simple" creatures or inanimate objects can't, because of the particular qualities of my subjective world-view.

It's not enough that my behaviour merely accord with some reason-based interpretation from the "outside" though. We want a theory that can illuminate the subjective reasons that are doing causal work on the inside. The upshot, it seems to me, is that in order to accurately individuate these subjective reasons we can't simply look at behaviour in the world. We have to look inside the head.

On Saturdays, if you see that I'm awake early, you could still *interpret* my behaviour as being reasonable, given that there are a number of objective mappings available to choose from in order to fit my actions to their inevitable outcomes. For example, one reasonable mapping is that I simply want to go to work on a Saturday. However, this is manifestly not *my* reason of choice. Such an 'interpretivist' programme goes about things precisely in the reverse order.

Philosophers of varied stripes have committed themselves to an interpretivist approach over the years because it is in accordance with the Naturalist Constraint (e.g. Dennett, 1987), but it does seem to violate the Cognitivist Constraint. The story, very briefly, is something like: we can't look inside the head for meaning because all we find are neurons and brute causal conditions, so look to what these neurons regularly and reliably cause instead (e.g. Davidson, 1975; Dretske, 1988). To act for a reason is just to act in accord with their behavioural outcomes.

On the face of it, this begs the very question we started with. We all know the commonsense causal relationships between mental states and their outcomes from our first-person experience. What we actually want is a theory that explains this, a theory that can predict what will happen *in virtue* of what a subjective reason truly is. Mere conformity to a reason's outcomes doesn't allow us to accurately retrofit that reason's contents into our heads. For example, my erroneous behaviour of waking up early on Saturday is in accordance with (at least) two different reasons' interpretations, because their functions are identical — they both result in going to work on a Saturday: 1) 'I want to go to work on Friday and I believe it is Friday', and 2) 'I want to go to work on Saturday and I believe it is Saturday'.

But one can't perform act-*Y* for reason-*X* if *X* is in fact *not* a reason for *Y*. If I know that it is Saturday, then my going to work cannot be because I desire to go to work on *Friday*. Something in our theory needs to adjudicate this problem, to say something about when and why I'm in error. By focusing on function, i.e. the fact that I am going to work on a Saturday, we instead resort to retrofitting the contents to fit the function it rationalises. Such a move deflates the specialness of reason-driven behaviour by time-travelling reasons backwards, after the fact. Reasons are supposedly special because they cause actions in virtue of how they make rational those actions. If we instead rationalise reasons in virtue of the actions they cause then we've given up the cognitivist constraint: all the causal work gets done prior, and without necessary appeal, to the reason-based interpretation of those causal states. Thus, any such appeal to reasons is rendered superfluous to explanation.

Now we've emphasised what we don't want, we can keep looking for an account that captures what we do want. The now pervasive account has been to assert that mental states and their brain-based vehicles must be *information bearing* in some sense. They carry information about the world in a form that I can store, process, retrieve, compare, update, and act upon. According to this notion, possessing such states is precisely what makes my behaviour seem purpose-driven and intelligent in a way that the behaviour of a rock or a planet is not. Rocks don't tumble down hills because they *desire* to be at rest, and planets don't orbit stars because they *intend* to follow the laws of gravity. They just do it. On the other hand, I do intend to get to work on time, and I follow the appropriate behavioural chain *on purpose* to reach that end. I do this because the intentions and beliefs about causal consequences of behaviour are information states circulating in my cognitive economy. Intelligent action is therefore special when considered as the lone class of causal process that results from *information processing*.

Appealing to information processing separates the intelligent from the brute physical world of rocks and planets in a neat way, by giving us some conceptual grip on the internal economies of behaving systems. Orbiting a star is quite complex, but we don't attribute merit to the planet for intentionally fulfilling that function; its orbit is *subject* to gravitational laws, not *sensitive* to them. Conversely, showing up to university early on a Saturday morning is quite stupid, but how and why I got there is still presumably (hopefully) worthy of attributions of intelligence. The causal mediation of the information bearing states in my head are what rocks and planets are missing. If this is the case, then what we mean by "intelligence" picks out something in the world regarding causal etiology more so than outcomes, and the notion of information processing gives us some explanatory purchase on that etiology.

Causal etiology and outcomes go hand-in-hand most of the time; one expects smart processes to produce smart outcomes. However, when it really comes down to adjudicating between the intelligent and the non-intelligent, it's the ability to do information processing that really counts. From the point of view of evolutionary science, successful outcomes may be what matter, but from the point of view of *cognitive* science, what goes on between your ears is what's special. There are "reasons" for the way the rock tumbles down the hill, and "reasons" for why the planet orbits the star, but they definitely aren't the rock's or the planet's reasons. In contrast, I definitely do act for reasons because I am not merely subject to the causal winds blowing about in the world. I am actually *sensitive* to them — because I have the appropriate information states circulating in my head. To paraphrase Fodor (2008), my conformity to certain reasons is partly explained by the fact that the *intention* to conform to them is an informational state in my head. I act for reason-*X* only when reason-*X* is the information state being currently processed/implemented (*ibid.* p. 38).

Information processing can thus give us a foothold in understanding the specialness of cognition, intelligence, and mental states, and do so in a naturalistic way. It putatively accounts for the causal etiology that is unique to intelligent systems by pinning down the *reasonable* way inputs can be mapped to a system's behavioural outputs. Prima facie, this also gets the individuation of reasons and how they're mapped to outcomes the right way 'round. It does this by appealing to the information contained within the system, and how that information is used. Changes to either the input, or the internal state that processes the input, will result in different outputs. The difference in outputs is reasonable though in virtue of the changes in informational contents therein.

This still requires more unpacking to satisfy the naturalists, but we're narrowing down the search: intelligent action is reason-driven, and acting for reasons can be plausibly naturalised using the concept of information processing. Specific inputs and outputs can be purely physical elements of the world, and the internal organisation of things like brains (or their non-neural equivalents) that mediate the transition from input to output can also be purely physical. But we need to flesh out those transition stages, whereby 'information' can be ascribed to the mediating states of interest. So, what is it that makes some arrangements of physical states, and not others, carry information content about other physical states? How can they be implemented in things like brains in a way that can be causally relevant and efficacious? The notion of 'representation' is popularly wheeled out here to do the heavy lifting.

Representations are familiar objects to us. Portraits, sculptures, scale models, road signs, and words and sentences are all examples of public representations available to us. They "represent" because they bring us into relation with something else, which is to say that they *stand-in*, or act as *proxies*, for other things. According to one influential story, with origins dating back to Charles Peirce (see Hardwick, 1977), this involves participating in a three-part relationship. The *representing vehicle* (the road sign, word, model, etc.) is some physical object, and the *represented object* is just whatever the representing vehicle is standing-in for. For example, the red and octagonal 'stop' sign is a representing vehicle whose represented object is the command to stop your car: the vehicle "carries" the object to you. The third spoke in the representing wheel is the effect of the representing vehicle on some user of that vehicle, with regard to the represented object. For example, I indeed stop my car when I'm confronted

with the stop sign. This representational effect has been variably termed the *interpretation* or *interpretant[1]*. On this *triadic* account (see Von Eckardt, 1993; O'Brien, 2015), 'representation' should be considered as a verb, rather than a noun; it is process that is only completed when all three spokes come into contact with one another. The "back end" of representation, so to speak, is the effect of interpretation, where the user of a representing vehicle comes into some contact with a represented object by way of that representing vehicle. The "front end" of representation explains why that is possible, and why that is *useful*, by implicating the vehicle's physical properties in certain relations with its represented object.

There are some important consequences of this approach, but first a possible regress must be confronted. In the case of public representations, the interpretation results in the production in the user of a representing vehicle about the represented object. Presumably, this is a *mental* representation in our heads that is then causally operative in how we selectively react to the original representing vehicle. But what is the interpretant of *that* mental representation? This requires some refinement if it is to pass the naturalist constraint unscathed, to avoid appealing to an infinite regress of mental representations. We can't simply invoke the antecedently representational if we are to explain in only material terms how it is that something becomes representational in the first place.

The focus must be placed on the representational *effect* involved with interpretation. Public representations are useful because they coordinate and regulate our behaviours with respect to their represented objects. The same can be said of internal mental representations, if we unpack the third spoke in the triad as the modification of behavioural dispositions towards the world (O'Brien, 2015; O'Brien & Opie, 2004). The infinite regress of representations is halted by bottoming out in the ultimate use of a mental representation towards some behavioural end. The capacity to use the properties by which a representing vehicle is in relation to its represented object, in a selective behaviour towards that object, requires no further appeals to representation. The naturalist constraint is also satisfied, because all three spokes in the triad can be cashed out in material terms. The represented objects are, presumably, some physical aspects of the world, presented to us in the physical form of sensory data, e.g. light hitting the retina. The representing vehicles are based in the neural substrate that make up our brains, and behavioural dispositions that result from our nervous systems are par for the course in the biological and neurosciences. To boot, this brings some precision to our concept of information processing, the special kind of causal etiology implicated in driving intelligent behaviours. Representation puts the '*inform*' in 'information', by giving an account of what kind of role a physical state must play in order to carry information about something else — it endows the user of that state with some relevant and selective alterations in behavioural dispositions.

There is a lot still to say about the way that different representing vehicles relate to their represented objects, e.g. the differences in how road *maps* and road *signs* relate to their represented objects. That discussion has flow on effects for understanding how different kinds of representing vehicle can be used to regulate behavioural dispositions (more on this in Chapter 2). For now, we at least have a broad outline of a theoretical hierarchy that captures our desired intuitions, wherein a modern cognitive

---

[1] I use these variations merely as a result of pedagogical bias, and I don't see that anything much turns on this. See Von Eckardt (1993) for one authoritative treatment that has slightly different nomenclature.

science can fill in the mechanistic blanks. To take stock, representation is a notion familiar to us, and on a triadic account is easy enough to transpose onto the domain of brain-based mental representations. Moreover, this account of representation lends rigour to the broad notion of information processing, where some meaningful criteria are laid out for what genuinely counts as an information bearing system. Information processing, in turn, gives us a theoretical grasp on the special kind of causal etiology that we think is fundamental to separating intelligent behaviour from other kinds of "crude-causal", or "brute-physical" processes. All of this, prima facie, keeps a privileged position in our world-view for intelligent behaviour and the explanatory frugality that comes along with it, without harming its specialness, and without harming our unity of naturalistic knowledge. All to the good, so far.

There remains one last piece to fit in to this theoretical landscape though, regarding the ordered transition of these representational states. The same physical properties that make them 1) useful proxies for objects in the world, must 2) ultimately shape their causal interactions, to truly fulfil the conceptual role required of intelligent processes. That is, we've sketched how representation makes sense of the notion of information *bearing*. But what about information *processing?*

1.3 Computation & the Mechanisms of Mind

The Computational Theory of Mind (CTM) was the *coup de grâce* that shepherded in our modern research enterprise called 'cognitive science'. The rise of CTM was the culmination of several lines of thought from distinct research areas, such as linguistics, neuroscience, computer science, and of course, philosophy. At its core is a methodological commitment to understanding the brain as a kind of computing device, and thus to studying the mind as in some computational relation to the representational states in the brain. In this way, CTM understands 'thought', or cognitive processing, as the sensibly ordered sequencing of such computational relations to representational states. For example, in one pathway of brain-computing medium, the rendering of a particular set of spatial representations in a particular order (among others) will be part of computing the function 'current visual scene'. The computational sequencing directs the representational states towards some problem domain because of what those representational states are about, i.e. how their content is related to that particular problem. It is in this way that computation can aid in the intelligent navigation of the environment: computation is a causal process shaped, in some respect, by the contents of the representational vehicles implicated therein (O'Brien & Opie, 2005).

The added appeal to computation over representational states is what separates *cognitive* science from the merely biological, physical, or neurological sciences. That is to say, the computational level offered an appropriate theoretical middle ground — in between the brute physical, and the vague mental — to postulate the material substrates of cognition and intelligence. The idea of computation finally gave some rich and physically tangible ground for cognitive scientists to begin probing for the mechanisms of mind and cognition, the culmination of centuries of philosophical refinement of the relevant theoretical landscape.

So far, this abstract idea of ordering representational states has been described in a way that is neutral on what *kind* of representational states are to be used. Computation is just a higher-order relation, a

causal sequence defined over — and driven by — the content of *some* type of representational states towards some task. However, that is not the way things have traditionally proceeded when defining computation. Fodor (1992) remarks that the '*big idea*' that set CTM off and running was the remarkable work of mathematician, Alan Turing. Turing developed a framework for the automation of symbol processing. As far as Fodor is concerned, this *big* idea of a symbol processor is the only *good* idea about the naturalisation of mind anybody has ever had (*ibid.* p. 6).

All one needs is a machine capable of discerning some discrete material properties of a physical symbol, such as weight, shape, or perhaps electrical conductance. The intrinsic properties of the symbol don't much matter, they can be any arbitrarily designated material, so long as the machine is capable of detecting at least two discrete states in that material. In recognising a true symbol, this discerning machine then proceeds to change its state according to a set of well-defined rules, embedded in its physical architecture, that dictate which specific changes to make in response to which specific symbols. By organising a particularly clever set of rules into the architecture of this device, and then feeding it a well formed string of symbols that it can discern, one can endow this device with a *semantically evaluable* capability: it can perform a computation over those symbols such that the resulting output is interpretable as respecting whatever that symbol string is intended to be about. The symbols are representational vehicles that bring their user, the device, into contact with some represented object(s), their contents, via their computational sequencing.

People are fond of remarking at this point that "the rest" is, apparently, "history". Digital computation was born, and the devices that recognise symbols—using high or low electrical conductances—got smaller, faster, more complex in their sets of rules and organisation, and became ubiquitous. It's no wonder that '*digital* computation' became synonymous with 'computation' *simpliciter*. However, this clearly shoehorns one specific kind of physical implementation of a computing device into the very notion of computation itself. We shouldn't conflate the engineering sense with the philosophical sense: what it takes to actually *build* a type of computer does not tell you *why* what you have built is, in fact, a computer.

The consequences for cognitive science have been to similarly shoehorn digital computation as definitional into its theorising about computational explanations of mental processes: the mind subsists on cognitive processes realised as neurally-based symbol manipulations. The end result is a framework that understands the mind as exploiting a kind of "language of thought", an approach developed to apogee in the eponymous book, Fodor's (1975) 'The Language of Thought'. According to the Language of Thought (LOT) hypothesis, cognition requires language-like symbolic vehicles of representation that are discernible and transformed by digital computational brains. They are transformed solely based on their material, or *syntactical*, properties, and the neural rules that constitute the syntax for transforming these mental symbols have been developed over evolutionary time. In this way, we have a naturalised notion for how mental states can be about something else, and cause behaviours because of those specific contents: their semantically evaluable content (which satisfies the cognitivist constraint) is respected by their syntactic causal liaisons (which satisfies the naturalist constraint).

Declaring that "the rest *is* history" inclines one to ignore the rest *of* history, however. Pre-dating Turing's automated symbol processor is a slew of remarkable devices used by our forebears to solve problems. These devices get completely overshadowed in contemporary discussions of computation, dwarfed by the omnipotence of Turing computation. Computation as an *idea* precedes the development of a symbolic, digital *implementation* of a computer by a long margin though. Although they were perhaps not titled as such, early problem-solving devices are computational devices too — especially when considered under the more fundamental, and implementationally *neutral* definition offered initially. These old devices use not only a different physical medium, but an alternative medium of representation. If they can do the work that computational systems were originally invoked to do, then there is no principled reason to ignore them in contemporary discussions of information processing, and in particular *biological* information processing. In modern computational terms, we'd call some of these devices *analog* computers.

Analog computers rely on measurable physical quantities whose intrinsic causal dynamics are governed by natural laws that are mathematically the same (or sufficiently similar) to those in the target domain being explored (Smith & Wood, 1959: cited in O'Brien & Opie, 2008: p. 60). In other words, the basic parts of analog computers relate to each other in ways that are comparable to the interactions and relations between the parts of the system they're being used to describe or model. The relations between the parts in an analog computer therefore *resemble* those of their problem domain, making it a literal *analog* of that domain. Our ancestors were ingenious developers of such devices.

A sundial, for example, exploits the similarity of the relationships between 1) time and the passing of the sun across the sky, and 2) the passing of the sun across the sky with the casting of shadows onto a circular plate. The physical mechanics that constrain the varying i) length and ii) position of the shadow cast on the circular dial's face are such that variations in the shadow length and position are sensitive to the linear ordering of moments in time. This is because the linear passing of time has a very obvious relationship to the cyclical movements of the sun across the sky. Even if a particular sundial is rather unsophisticated, and is a poor indicator of more modern and formalised units of time, e.g. minutes, the length and position of the shadow are nonetheless sensitive to important periods or moments of the day/night cycle. For example, midday corresponds to the shortest shadows cast, and dawn/dusk with the longest shadows cast, so it is nevertheless possible to see *roughly* where in the day/night cycle you currently are.

Exploiting physically comparable and covarying features of the world inevitably became more refined. Astrolabes and sextants, for example, were developed by pioneering navigators and astronomers to determine directions, times, and swathes of other celestially-derived information. These devices also work by exploiting specific geometrical relationships that hold between 1) three-dimensional spaces and objects in the world, and 2) some conserved counterpart relations that exist between the various parts of the devices. For example, the astrolabe exploits various features of geometry, like distance and angular deviation, and how they (or their ratios) can be combined and conserved across scales with fewer dimensions. For now, an easier example of this kind of conservation of geometrical properties is the orrery.

An orrery is a functioning mechanical model of celestial bodies, demonstrating their positions and movements relative to each other. Typically, they are models of our solar system, either with the sun at its centre (heliocentric), or with the earth at its centre (geocentric). Orreries usually use spheres to represent the planets or sun, naturally. These spheres are then attached to arms of various length that connect ultimately to a central pillar, and are distributed about the central sphere in their correct relative order. The arms are, in turn, attached via the central pillar to sets of cleverly organised gears which govern their rotational behaviour about the central sphere. The design of the interlocking gears is such that the spheres' orbits about the central sphere, relative to one other, accurately represent the orbits of the planets around our sun.

It should be obvious that these early mechanical devices, like sundials or orreries, are not *symbol processors*, because symbols are arbitrarily related to their targets, and do not get their contents through a similarity relation. Symbols only become semantically evaluable after being disciplined by the causal infrastructure of the computing device they're fed into. Yet the varying shadows on a sundial, and the rotating spheres in an orrery, certainly fit the broader and more theoretically fundamental definition of computers that was set out above: they are a set of representing vehicles which can be exploited by some user-system in order to come into some contact with another domain. As a result of that contact, they facilitate the solution of some problem(s) with regard to that domain, in virtue of their specific representational contents. The representing vehicles in an orrery are superficially the spheres, but in some important sense the gears that drive the spheres must also be given consideration[2]. In virtue of the spheres' relative positions and orbits about the central sphere, they depict the solar system as being a certain way: their *representational contents*. The spheres can then bring some user into contact with a separate domain of represented objects: the positions and motions of the actual planets in our solar system. However, an orrery's spheres do this in a non-arbitrary way, because their positions and movements in the model orrery — relative to one another — *resemble* those same relations between the real planets. There exists a similarity relation between the representing and represented domains.

If the orrery is designed well, it can then be used to compute a number of problems. First, it can tell you superficial things like the planets' order of distribution, ranging from closest to farthest from the sun, and their directions of orbit (calibrated from some astronomical vantage point). More interestingly though, it can aid in the solution of problems to do with prediction. For example, if the orrery is set up to accurately show the current positions of the planets, and conserve their relative speeds, it can be then used to predict where the planets will be relative to each other at some future time, by simply cranking the model forward. Instead of having to merely *wait*, one can look forward to a distant time, but only so long as the model preserves all of the planets' motions well enough. The positions of the planets can then be used to make predictions. For example, an orrery can assist in predicting which planets will be visible at some time from some point on the earth (which is useful for navigation), the moon cycles (for calendar keeping), planetary alignments, eclipses, and so on. However, *why* it is possible for an orrery to be used accurately for prediction is illustrative of a vital difference between

---

[2] The spheres are essentially cosmetic aids — any other marker placed on the rotating gear could be used, to a greater or lesser cosmetic effect. But, at any snapshot in time, the relative positions of the markers (whatever they may be) on the gears are usually doing most of the representational work in an orrery — depending on the specific task at hand.

the two species of representational vehicles: sundials and orreries are *not* symbol processors. But if they are analog *computers*, what are these non-arbitrary representational vehicles they in fact implicate in their processing, in order to resemble their target?

The sets of gears in an orrery are made to scale in such a way as to preserve the relative positions of planets during their orbits, according to the idiosyncratic spacing of the planets in any given model orrery. All of the arms, and therefore all of the proxy planets, are moved by a common crankshaft (or automated equivalent), and so move at a common time interval. However, all of the real planets' orbits are vastly different in size, and completed at vastly different rates. If an orrery were to conserve the accurate ratios of distances between the planets, and therefore the relatively enormous circumference of some of the outer planets' orbits, it would quickly become unfeasible as a practical model — because the model itself would become enormous. To adjust for this problem, the distance between the orbiting arms of the orrery must be reduced, and a typical orrery has roughly uniform spacings between each orbit. This makes the entire model much easier to sight, and thus to use. However, the speeds of the outer arms must be radically lowered in that case, specifically in proportion to the total reduction in the size of their orbits. If you don't adjust their speeds their orbits will become useless as models of their respective planets' orbits, as they would be orbiting much too fast, and the model would no longer preserve the proper positions of the planets with respect to each other and the sun. Naturally, this would render the model useless for prediction. This is why the geometric properties of ratio or scale are central to the computational properties of the orrery.

Ratios or scales are not absolute or immutable quantities, because they don't relate to "first-order" properties of objects. First-order properties are superficial things like shape, colour, or weight. Rather, ratio and scale are "second-order" properties that arise from the relations *between* objects (see O'Brien & Opie, 2004; 2005). These second-order relations can therefore be conserved across radically different sets of objects, with radically different first-order properties; planets and brass spheres could hardly be more different in their first-order properties like material substance, size, and weight. However, while the absolute distances between the planets, and their actual speeds of orbit obviously cannot be preserved by the model, their *relative* distances and *relative* speeds can — by preserving their ratios at some set scale.

Imagine all of an orrery's planetary spheres aligned perfectly and in their correct order, but uniformly spaced. The ratios of their relative speeds could be shown by how much they all moved away from their original position of alignment after some chosen time interval, e.g. one model-Earth day. The Mars-proxy would move less than Earth's, as its orbit is 1.8x the size of Earth's. Venus and Mercury would move much more, on the other hand, as their orbits have much smaller circumferences and thus can be completed in a few Earth-months. In this way, the real planets' orbital speeds and positions relative to each other (a *second-order* relation) can be preserved by the orbiting spheres in an orrery: their relations resemble those of their target domain.

Preserving the relative movements of the larger planets' spheres accurately is dependent on the clever design of the gears though. The gears transfer movement at different rates from the central shaft out to the assorted compounds of planetary gears. They reduce or amplify the rotation depending on their comparative circumferences, and thus the ratios of teeth they have to each other. Doing this for an

inner planet is relatively easy; just make the gear smaller so it orbits faster than its neighbours, and the circumference of the gear will resemble (roughly) the elliptical shape of the real planet's relative orbit. But given that the outer spheres' orbits are drastically reduced, their larger gears must be turned even slower than their mere size would dictate. That is, some computational relations must be made different here, in how one gear causally relates to another; a reduction of speed must be computed to fit the new ratio of movement required for an outer sphere's smaller orbit, with respect to the inner spheres. Suffice to say that by placing an intermediating set of coupled gears together, into a 'double reduction gear', this can can be easily achieved[3]. The relative size of the two gears in this coupled set, with respect to each other, ultimately determines the exact amount of change in the transfer of movement between the initial gear and the desired end gear. In some important respect then, this coupled reduction gear also resembles the necessary change in ratio that it is being implicated in computing. It physically embodies that ratio change in its physical size because physical size and teeth ratios determine the reduction or amplification factor. Thus, the amount it changes the rate of rotation that it receives into the rate which it passes on to its adjacent gear is intrinsic to the geometric, and thus very much *physical-causal* properties of the gear.

The way that the spheres and gears in an orrery resemble their target domain's relational, second-order properties are therefore directly determined by the geometric properties instantiated by some aspects of their physical structure. It is important to note that all of the causal work inside an orrery is being done by that same structure. The elements in the system that are being used to stand-in for something else are doing double-duty here: the *representational* properties are one and the same with the *computational* properties of the system as well.

In a digital computer, there is a fundamental division of labour between the hardware that instantiates the most basic kind of "machine rules", and the symbolic representations to be acted upon by that hardware. This distinction is collapsed in the case of the orrery and its planetary proxies. The representing vehicles to be sequenced computationally are also doing the computational sequencing themselves. Indeed, this is because they *are* the hardware. The second-order resemblance relation between an orrery and the planets only holds because of the very real, very physical, *structural* properties of the relevant elements within the systems. However, it's not just the mere size of the gear that determines what it represents and how it computes. What matters is the gear's size *with respect to* its complementary place in *a set of* variably sized gears, and the geometric relations they manifest as a result. Exploiting non-arbitrary resemblance relations, where some of these geometrical properties of a represented domain are replicated in some representing system, has been called '*structural representation*' (see O'Brien & Opie, 2004; 2005).

This is a crucial distinction between digital and analog computers. They use fundamentally different species of representations that physically encode information in a very different way — one symbolic, the other distinctly non-symbolic. This has significant consequences for how the representations are to be processed in a physically implemented computing device, a distinction that is often ignored. Symbol processors are different mechanisms altogether, and thus require fundamentally different physical infrastructure to their analog counterparts that exploit structural representations. We are

---

[3] See https://en.wikipedia.org/wiki/Gear_train#/media/File:AnimatedGears.gif

therefore presented with a fork in the road: which account of computation is the better theory to explain intelligence, and the relationship between the mind and the brain? Is the brain a digital computer, shuffling arbitrary symbols according to formal rules embedded in its architecture? Alternatively, is it an analog computer, where its causal cogs structurally resemble some relations and properties of their target domain, much like an orrery? These are fundamentally different computational mechanisms, and something needs to arbitrate between them. One attempt at arbitrating between these varied approaches is the focus of the rest of this discussion.

In the following chapters, we'll depart from our theoretical landscape that we have been refining (see *Fig.* **1**) and dive headlong into some empirical investigations of the world. This shift in approach is so that we can discover something about the real mechanisms that ostensibly underpin the minds and the kind of intelligence we're most familiar with — biological intelligence. The problems being discussed, until now, have been distinctly philosophical in nature. We have considered some concepts, and the conditions for what we think it would take to be an intelligent system: what makes it possible for some system to act reasonably according to the way that it takes the world to be? But these philosophical considerations have motivated more practical questions, as we have begun to encounter problems to do with different physical implementations of those concepts and conditions. Of course, this is not a new dilemma.

C. Randy Gallistel is one contemporary and influential cognitive scientist who has argued exhaustively in favour of the digital computational theory. Gallistel is a staunch computationalist when it comes to cognitive science (see Gallistel & King, 2010). As such, I think it is fair to guess that Gallistel would agree on the philosophical landscape, and the motivations, I've set out above.

We disagree, however, on the questions regarding engineering. It is my view that Gallistel falls victim to the problem of shoehorning *digital* computation into the primitive notion of computation itself. This is not a trivial error, as much turns on this. As a result, his *computationalism*, i.e. a philosophical position with regard to the role of information processing in explaining minds and intelligence, becomes a parochial kind of *digitalism*, i.e. an empirical position on the engineering and physical implementation of a computationalist philosophy.

Gallistel argues that mental representations must be symbolic expressions, composed of atomic symbols with semantic and syntactic properties, just like the symbol strings fed into a digital Turing computer. These symbolic expressions are manipulated by the formal rules embodied in the computational hardware of the brain. One of the most illustrative examples Gallistel uses to substantiate his arguments is by analysing the curious case of the Desert Ant and its foraging behaviour. The Desert Ant demonstrates a remarkable ability to navigate a harsh environment that is absent of landmarks during foraging behaviour. From a theoretical perspective, their navigational aptitude is so illustrative because it epitomises the phenomenon that we're interested in: what does it take for a system to act reasonably and flexibly given the way it must take the world to be?

The ants must hold some kind of internal representational states regarding their movements and their changing position relative to their nest, and be able to do computational work on those representations in order to integrate their outbound paths and return home once they have found a food source. They
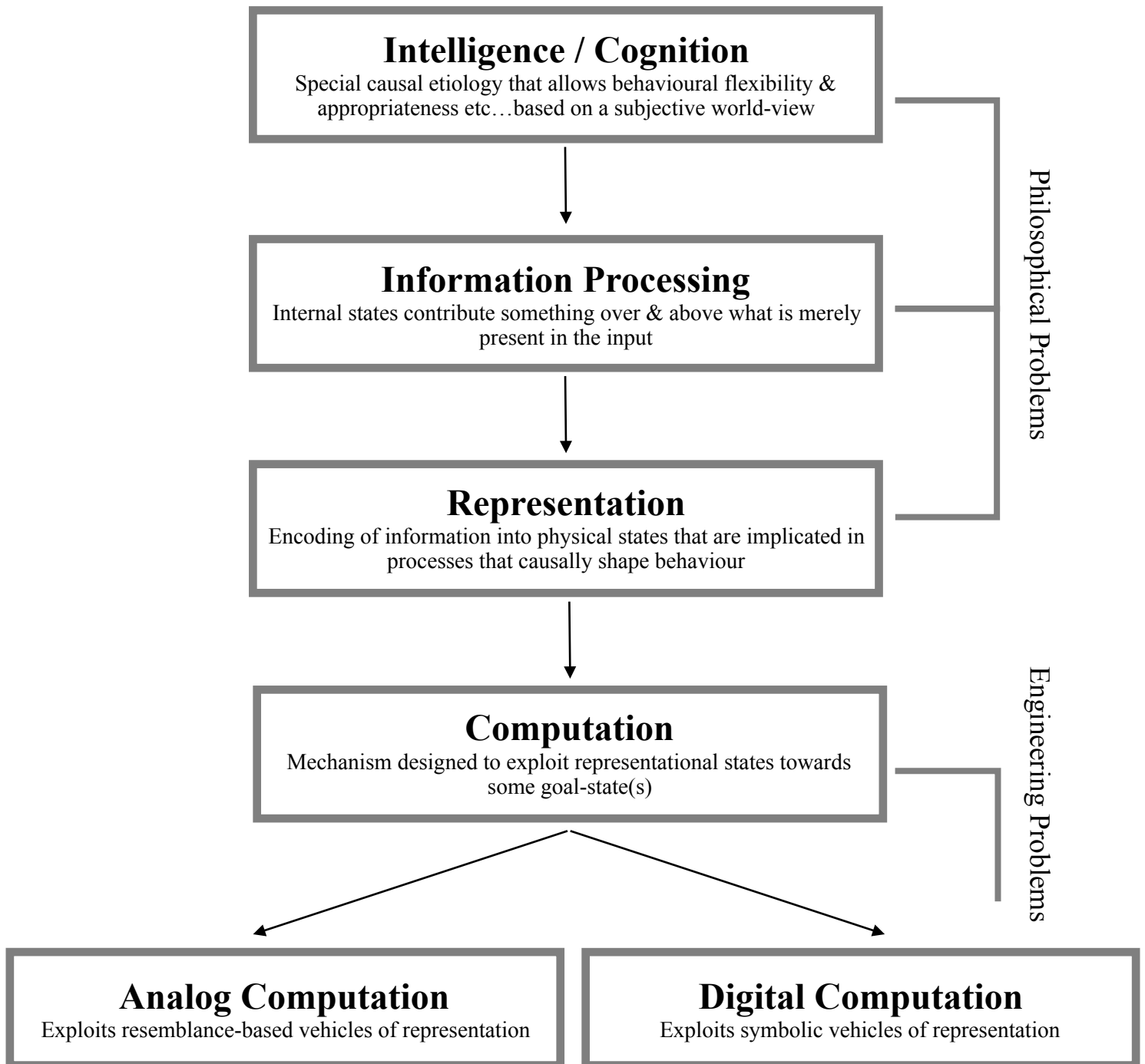
Intelligence / Cognition
Special causal etiology that allows behavioural flexibility &
appropriateness etc…based on a subjective world-view

Information Processing
Internal states contribute something over & above what is merely
present in the input

Representation
Encoding of information into physical states that are implicated in
processes that causally shape behaviour

Computation
Mechanism designed to exploit representational states towards
some goal-state(s)

Analog Computation
Exploits resemblance-based vehicles of representation

Digital Computation
Exploits symbolic vehicles of representation

Philosophical Problems

Engineering Problems

Fig. **1:** The theoretical hierarchy at the heart of cognitive science, and of this discussion. Each box represents a stage of conceptual refinement in the philosophical landscape, and the core postulate or discovery that cemented its place there. Each lower postulate is necessary to theoretically substantiate, or physically implement, its predecessor. The chain of reasoning is further sub-divided into two different categories of problems: the philosophical, and the engineering.

therefore take the world to be a certain way, and that rendering of the world includes *something* about their geographical position in it.

From an engineering perspective, Desert Ant navigation is a great case study because ants are relatively simple creatures with small and comparatively easy-to-analyse brains. Secondly, the computational task under examination is very circumscribed and thus easy to define clearly: "path integration", the process of navigating without landmarks, is both felicitous to mathematical formalisation and very well understood. By Gallistel's lights, the kind of environmental variables available for the ant to integrate with, and the kind of computational operations they must perform over data extracted from those variables in order to return home, necessarily requires *symbolic* representations of those variables and thus a *digital* computational apparatus in their brains.

Thus, the Desert Ant presents a perfect encapsulation of the problems at the heart of the philosophy of cognitive science. It has philosophical importance for understanding intelligence, representations, and computationalism. It also offers up a boots-on-the-ground test case for our various accounts of computational devices that require different engineering altogether, so we can test our theoretical commitments against the worldly data. I share Gallistel's broad concerns regarding the philosophical commitments of representation and computationalism underlying biological intelligence and behaviour. In the case of Desert Ant navigation, however, I will argue that Gallistel's analysis of this phenomenon is incomplete in two respects which lead his conclusions astray.

Firstly, Gallistel does not consider the possibility of analog computers, with their concomitant structural representations, as physically implementable information processors. This is because Gallistel (Gallistel & King, 2010: p. *viii*) is wedded to the idea that, "…there must be an addressable read/write memory mechanism in brains that encodes information received by the brain into symbols (writes), locates the information when needed (addresses), and transports it to computational machinery that makes productive use of the information (reads)".

This narrow focus on digital computation necessarily affects the way his theoretical analysis of Desert Ant navigation bleeds into his beliefs about how to engineer a solution to the problem. As a consequence, Gallistel naturally makes parochial demands of *any* potential explanation of the phenomenon, i.e. that any explanation of Desert Ant navigation *necessarily requires* a symbol processing mechanism at its core. However, as introduced above, information processing is not only achievable digitally, over symbolic expressions. There *is* a different framework that is at least worthy of consideration.

Analog computers, such as orreries, can be used to carry and operate over representations, and thus compute information, but can do so without the need for *symbolic* representations. One theoretical framework of analog computation that is neurobiologically-inspired has been called '*connectionism*'. In the course of arguing that the Desert Ant brain needs to instantiate a symbolic processor of a particular kind, Gallistel takes aim at, and mischaracterises (according to me), the connectionist framework. Gallistel considers connectionism as just one way of physically implementing a digital computing device, thus construing connectionism as a symbol processing framework. When he analyses the practical application of "connectionist models" (under his characterisation) to the problem

of Desert Ant navigation, he subsequently finds that they lack the power to implement the particular kind of symbolic processor that he thinks is necessary to complete the task: a read/write memory. Connectionism, therefore, cannot even explain simple cognitive tasks, according to Gallistel, and should be relegated to the rubbish heap of theories of cognition.

However, there exists a respectable tradition that considers connectionism as a fundamentally distinct computational paradigm, *not* as a subset of digital computational theory (Rumelhart & McClelland, 1986; O'Brien & Opie, 2005; 2008). Much debate has surrounded this issue of the theoretical commitments of connectionism, and whether it indeed differs from the 'classical' digital computational story (see Fodor & Pylyshyn, 1988; Churchland & Sejnowski, 1990; Horgan, 1997; Horgan & Tienson, 1992; Churchland, 1998; Bechtel & Abrahamsen, 2002)[4]. Without getting too lost in that debate, I take the position that the strongest and most reasonable interpretation of connectionism is an analog computational one. Consequently, by my lights, Gallistel constructs a straw-man of connectionism by assessing connectionism's explanatory credentials, with specific regard to the problem of Desert Ant navigation, with measures that are actually the special preserve of digital computers. In virtue of this mistake, i.e. mismeasuring connectionism's capacities, the way in which Gallistel dismisses connectionism's explanatory credentials as a research paradigm in the cognitive sciences *tout court* can be considered premature, to say the least.

Secondly, Gallistel's empirical analysis of the navigational task to be solved by ants does not appreciate the phenomenon in enough empirical detail. In particular, Gallistel fails to attend to the neurobiology of the ant and how their neurobiology relates to the geophysics of their environmental circumstance properly. I will argue that, after looking more rigorously at the empirical data and the nitty-gritty of the relevant environmental variables, it becomes apparent that Gallistel's depiction of the problem to be solved by the Desert Ant is in fact incorrect. Combine this with his narrow framing of how any computational mechanism capable of solving navigation problems are to be engineered, and Gallistel makes demands that are therefore both unnecessarily computationally parochial as well as incompatible with the relevant empirical data. I ultimately will argue that the alternative account of computation that is on offer, i.e. analog computation and the structural representations that come with it, can mesh perfectly with the fuller depiction of the ant neurobiology and the environmental task it is implicated in solving.

Fodor (1987, p. 97) writes: "I want a naturalized theory of meaning; a theory that articulates, in nonsemantic and nonintentional terms, sufficient conditions for one bit of the world to be about... another bit". Gallistel has answered this call in one way, but I prefer another. Using the illustrative example of Desert Ant navigation, the primary argumentative burden of this thesis is to untangle and explain the differences in our accounts. Ultimately, I will argue that an analog computational framework fits the data regarding Desert Ant navigation more closely than its digital competitor. It does so while retaining all of the philosophical commitments that are necessary for a system to be a representing, information processing, and intelligent system.

---

[4] This debate is, in fact, enjoying somewhat of a renaissance, with renewed interest in the viability of structural representation and its explanatory credentials being tangible in the literature (see Bartels, 2006; Isaac, 2013; Beck, 2014; Morgan, 2014; Azhar, 2016; Boone & Piccinini, 2016; Nestor, 2016; Maley, 2017; Williams, 2017; Williams & Colling, 2017; Thomson & Piccinini, 2018). The "representation wars" (see Clark, 2015), it seems, are as violent as ever.

1.4 The Project & Path Ahead

The remainder of the discussion will contain two arguments running in parallel; one argument is between analog (connectionist) and digital (classical) computation, the other between myself and cognitive scientist, C. R. Gallistel. My dialogue with Gallistel over the explanatory credentials of connectionism, via the case study of the Desert Ant, is a fine-grained proxy for the larger philosophical debate regarding the theoretical foundations of cognitive science, and cognitive explanations in general. I will argue that several mischaracterisations permeate Gallistel's arguments against connectionism. They follow a similar structure to the outline above, where one component is an abstract analysis of theoretical or philosophical commitments, and the other component concerns the practical implications of those philosophical analyses.

So, there will be some interwoven threads of argumentation to keep track of. The debate between classicism and connectionism in the philosophy of cognitive science is the broadest and overarching theme: what is the "true" relationship between the mind and brain, and of information processing and intelligence? Is the mind/brain much like a digital computer, shuffling symbols according to their syntactic properties? Alternatively, is it more like an analog computer, exploiting certain physical similarities that exist between its architecture and that of the external world? Chapter 2 will bring out the philosophical and engineering differences between these two accounts in some more detail, focusing on the independence and uniqueness of the analog computational framework.

The significance of how we ultimately answer the above questions for our understanding of the natural world, and our place in it, hardly needs repeating. The difficulty in adjudicating between different answers to these profound questions, however, *does* need repeating — many people are unaware that there even is an issue to be adjudicated. The proof of the pudding, I say, will be in the eating. It's my view that, at least in *mainstream* cognitive and neuroscience circles, the connectionist option isn't even on the menu. If it is, it usually has the wrong ingredients listed.

By going through a particular case study in Chapter 3, i.e. examining how the humble Desert Ant and its morphological cousins navigate, these knotted arguments to do with understanding cognitive and computational systems can be untangled. It so happens that this is one example where I believe a prominent theorist has done connectionism an injustice, and so represents a perfect point of contact between the competing theories. If our theories can survive collision with empirical data, the epistemic gold-standard under naturalism, then we can continue scaffolding up from there. So we will start small, and Gallistel and I agree at least on this point: if we can't explain some of the "simplest" cases of cognitive phenomena with the theoretical tools at hand then we may need to fashion ourselves some new ones. Before we invent some radical new tools, as Gallistel (among others) has turned to doing with "molecular computation" (see Gallistel & King, 2010), we should first give the standard framework a fair go.

The Desert Ant case study involves some technically detailed empirical puzzle-pieces that will be very much in the foreground for the core of the discussion. There'll be crash courses on concepts from astronomy, geometry, and neurobiology. All for good reason though. I hope to convince the reader that it is only by looking in detail at how the mechanisms that mediate ostensibly intelligent behaviour

relate to their environment, warts and all, that we can find clues for naturalising cognition — and thus understand the philosophical and mechanistic foundations of intelligence. Most importantly, my central thesis is that the independence and true *attraction* of connectionism as a theoretical framework can be demonstrated and fully appreciated as a result of such a rigorous analysis. The devil really is in the details.

Connectionism can approach and explain cognitive phenomena in a unique manner. This uniqueness and independence is lost under other more common interpretations of connectionism, i.e. the 'implementationalist' interpretation of connectionism, which sees no real philosophical daylight between it and classical computation. Gestures will be made towards the added explanatory purchase that cognitive science can gain in adopting the "proper" connectionist framework, and how future avenues of research open up in light of doing so. However, these will just be gestures, and are not the central point. The reader may maintain that, after all, the classical approach to computational cognitive science is still superior. I'm under no illusion that I will settle the debate once and for all[5]. Connectionism is due a fair hearing though, and it is my position here that the strongest possible form of connectionism is commonly elided, the argumentative space filled by a straw-man instead. I accuse Gallistel of doing just this, and Chapter 4 will outline why I believe so. Thus, my main intent is to provide an illustrative defence of the unique approach that connectionism offers to our investigations of cognition and its place in the natural world, when connectionism is properly understood as an analog computational explanatory framework.

In Chapters 5 and 6 I argue that analog computation is in the driver's seat of the brains of Desert Ants. Which is to say: I believe that the underlying computational hardware that guides their path integrating behaviour is best understood as being analog. However, *why* such analog computational mechanisms are so *successful* at guiding this behaviour, by my lights, is the interesting *philosophical* point to take home: the analog computational framework can offer a representationally rich, causally efficacious, and entirely naturalisable metaphysics of mind, cognition, and intelligence. To boot, the mathematical elegance and simplicity in how an analog computational explanation sheds new light on this particular problem is aesthetically pleasing. Ultimately, the weakest conclusion to take home is that I *at least* intend to show that the demise of connectionism is not a foregone conclusion. Contra the claims made by Gallistel, I will argue that connectionism remains a viable and robust research programme in the cognitive sciences.

---

[5] You can bring a horse to water…

Chapter 2

Computation & the Philosophical Foundations of Cognitive Science

Rival Approaches to the 'Big Idea'

"[It] is beneath the dignity of excellent men to waste their time in calculation, when any peasant could do the work just as accurately with the aid of a machine."

- *Gottfried Leibniz*

2.1 Cognitive Science: The Big Idea

The notion of information processing is the lodestone of cognitive science. It is the conceptual distinction that cleaved *cognitive* science apart from its related fields of inquiry, like behavioural neuroscience or biophysics. It forms the basis of a fully independent research methodology with a unique domain of inquiry. Packed into information processing is traditionally, under the guiding light of naturalism, a commitment to both the representational and computational theory of mind. Taken jointly, these commitments are widely recognised as the means by which organisms identify and adapt to changes in their circumstances. Via the computational theory of cognition, we're equipped to understand mechanistic explanations of these capacities.

The added explanatory power of a computational level of analysis bridges the gap between the undoubtedly relevant, but merely causal explanations of behavioral neuroscience, and the intentional (i.e. *intelligent*) explanations we seek to naturalise. Chapter 1 discussed why, historically and philosophically, this bridge is widely accepted as necessary to make (see *Fig.* **1**). This chapter will turn to looking at differences in the kinds of computational accounts on offer in more detail. This involves demonstrating how these accounts posit different physical mechanisms that can be engineering solutions, the physical bridges, to the kinds of philosophical gaps we think need to be crossed in order to naturalistically account for cognition and intelligent behaviour. The argumentative burden of this chapter will be to demonstrate that the way we conceive of 'representations', and *how* they get to be information bearing, will have significant consequences for how we then apply that representational theorising in a physically implemented computing device.

To briefly reformulate the problem as set out in Chapter 1, the way that the cognitive science literature sees 'computation' as the nexus between the engineering and the philosophy of intelligent systems is that it offers a physically implementable solution to the *content causation problem*. O'Brien (2015: p. 2) gives one formulation of the content causation problem, which I've modified below:

> 1) Intelligent processes are causally efficacious of behaviour in virtue of the representational contents implicated therein.

> 2) The representational contents of intelligent processes are determined by the intrinsic properties of the brain.

> 3) The brain is causally efficacious of behaviour in virtue of its intrinsic properties.

The first statement encapsulates the first three steps in the theoretical hierarchy from Chapter 1 (see *Fig.* 1), i.e. that intelligent and cognitive processes are fundamentally different from other physical processes in the universe. It reiterates the Cognitivist Constraint — that the way in which intelligence is different from other physical processes is that it is a matter of selective identification and response to variables in one's environmental circumstance. This is enabled by the rendering of a subjective world-view, i.e. the *contents* of an internal, representational model.

The third statement encapsulates all that we know about standard behavioural neuroscience, and reiterates the naturalist constraint. We know that the brain is the arbiter of our behavioural responses, and, for example, how circumscribed losses to modules of brain-matter bring only circumscribed deficits in cognitive capacities. The intrinsic properties of the brain are therefore determining the complex set of causal relations we enter into with the world.

The second statement is the all-important nexus between the first and third, wherein their apparent insolubility, and the *over*-determination of behaviour can be addressed. It is the prosaic yet controversial thesis that, at least some of the time, the intrinsic properties of the brain are both content determining and behaviour determining. Thus, intelligent processes can be metaphysically robust in the sense that they can truly be information bearing, while the naturalist and cognitivist constraint can also be reconciled: representational content can be a naturalisable property of brain states, *as well as* causally efficacious of behaviour.

The next sections will bring out just what properties of the brain we need to pick out, and exactly what kinds of relations they need to enter into in order to be content determining. This will demonstrate why it is known as the content causation *problem*, and why the second statement is seen in some corners as controversial, because much turns on this simple proposition: not all types of representations are equal. Different representations require altogether different content determining relations. These content determining relations, in turn, require different properties of their representing medium to subsist on. Hence, *what kinds of representations* we posit as the brain-based medium for computation will determine how we go about solving the content causation problem. By implication, this will determine just what kind of "cognitive science" we end up with.

These are consequential differences that will be emphasised in distinguishing between the two computational theories on offer in the discussion to follow. It will be central to the arbitration process, to assess which theory (if either) can survive collision with both our philosophical commitments *and* empirical findings. In sum, the major task of the rest of Chapter 2 is to answer this question: is there more than one way to make rigorous and precise the notions of representation and computation — in order that they may play a robust explanatory role in modern cognitive science?

2.2 Rival Approaches I: Classicism

The next sections will plot the general forms and a potted history of the competing computational theories on offer: digital vs. analog. In later sections the details of their specific theoretical commitments, and engineering consequences—with specific regards to the problem of navigation— will be filled in. To begin, the 'classical' approach to computational theorising will be outlined.

Jerry Fodor (1975: p. 175) begins his book by quoting President Lyndon B. Johnson's line that, "I'm the only President you've got". What he is alluding to in his introductory chapter is, of course, that he believes that some form of the *Language of Thought Hypothesis* is the "only game in town" when it comes to solving the content causation problem (*ibid.*). The LOT hypothesis posits one particular representational medium for brain-based computation: a language-like symbolic medium which is composed of atomic units that can be permuted into cognitive "sentences" — a kind of *"mentalese"*.

The LOT asserts a parallelism between the power and flexibility of linguistic expressions, and those of cognitive operations. We learn a language by acquiring some atomic vocabulary, along with sets of instructions. The instructions inform how we concatenate atoms into well-formed strings of atoms, molecular constructions, according to that language. The sentences that emerge from this piecemeal process then *mean* — i.e. represent — whatever the contents of the individual atomic units mean together with their specific relational structure; their ordering in a sentence. For example, 'Dog loves Cat' means something different to 'Cat loves Dog' even though their atomic units are identical, because the subject-verb-object ordering of the English language makes it so. Such representing systems allow for indefinite growth in complexity, and limitless representational flexibility in their constructions. An example for the cognitive domain may look like:

$$(LIFT)\ x = ((MOVE)\ x) + ((MOVE)\ up)$$

The complex concept of 'lifting' some object x is a function of much "simpler" concepts like 'causing movement', plus the particular 'direction of movement' that is desired (adapted from Piantadosi, 2016: p. 4).

As a contemporary and influential theorist of the Fodorian-stripe, C. R. Gallistel calls this capacity for virtually endless concatenation of atomic symbol structures into molecular symbol structures the "essence of compositionality" (Campbell, 2015). Compositionality is a notion taken from Fodor and Pylyshyn (1988), and, according to Gallistel, compositionality is at the core of computation. In order to demonstrate this point and really get a grasp on these fundamental principles, Gallistel strips the LOT hypothesis down to a less abstract level than that of cognitive concepts, and applies the same fundamental logic to a more "basic" class of problem: simple arithmetic.

To physically realise simple arithmetical operations there are two requirements: 1) physically realised symbols to represent real-values, i.e. numerals to represent "numbers", and 2) physically realised machinery that is capable of combining the symbols to produce arithmetically respectable outputs. This is no different to the fundamental principles of the LOT, or to the learning of a natural language — there must be some atomic units of meaning, and then some rules according to which molecular units can be constructed out of those atomic units and then manipulated.

Computer science has a standard approach to engineering such symbol processing systems capable of performing arithmetic. A symbol can be anything, as long as it has an adequate margin of error in its discernible syntactic properties. This is important so that one symbol can't be confused for another symbol, and reduces the effects of noise in the signal. High and low voltage states perform this function perfectly, as once a voltage state reaches a certain threshold it can be considered "high" enough to "open" a respective electronic "gate". The gate is thus entered into its active "1" state, where its output can now be used by other gates down the line. Anything below the voltage threshold and the voltage gate will remain quiescent and closed, thus representing a "0" state. From these two clearly distinguishable *binary* symbols, some more abstract machinery can be built.

The binary nature of truth (in classical logic, at least) makes it possible to exploit a system of logical argument concerned with binary values, true or false, in operating over a system of binary arithmetic, and this system is called "Boolean Logic" or "Boolean Algebra". The logic gates under discussion are

therefore usually called "Boolean Logic Gates". Boolean logic can be used to analyse and link together simple logic gates to implement more complex logical operations[6]. The simple voltage gate described above merely indicates whether or not its input is present — "1" for true, "0" for false. The inverse of this gate is called the NOT-gate. The NOT-gate output is "1" or true so long as its input is "0" or false, and vice versa. This is equivalent to the operation of logical negation. Not much can be done with these two operations, however.

Building from these single-input fundamentals, more sophisticated electronic circuitry can be designed so that, for example, a logic-gate is opened only if *all separate* input lines into the gate are *all* simultaneously activating above threshold. The logical function such a gate implements is the AND-function in Boolean logic. This function is formally equivalent to a logical conjunction: a conjunction is true if and only if all operands in the conjunction are true. A simple 2-input AND-gate is commonly represented as in *Fig. 2*.

If we construct a truth-table for the above AND-gate, the overlap between the binary nature of *logical* truths and binary *arithmetic* truths becomes more apparent (see *Fig. 3*). The table shows the output the AND-gate gives in response to any one of its four possible combinations of input. Note how the above table would be identical to a table representing the basic arithmetic operations of *multiplying* 1 and 0.

The same four combinations of variables are possible, and they produce the same column of output variables. Zero multiplied by zero is zero; one multiplied by zero is zero; and so on. Another of the fundamental Boolean-gates is the XOR-gate, otherwise known as the "exclusive-or". The formal equivalent that the XOR-gate implements is an "exclusive disjunction" (see *Fig. 4*). Note again how the truth-table for XOR-gates (*Fig. 5*) looks (mostly) like what a truth-table for *adding* one and zero together would look like[7].

Using this system, one can encode any number using a *N*-bit binary string, and then perform any number of logical operations over them with enough of the right kinds of Boolean-gates. As a result, it is possible to show something true of their arithmetic interpretation using only formal, Boolean operations. For example, '0001' and '1011' encode the numbers '1' and '11' respectively in *4*-bit binary code. With some clever circuitry using the Boolean-gates above, one can perform any of the arithmetic operations over those 4-bit strings. Forgetting some unnecessary details for the moment, a "half-adder" circuit (see *Fig. 6*) completes half of the process of addition. A half-adder does this with only a single AND-gate together with a single XOR-gate. Naturally, putting two half-adder circuits together into one complete unit of abstraction then gives one a "full-adder" circuit.

Now we can return to Gallistel's point about compositionality being at the heart of computation. Using this classical approach that combines binary encoding for representation with logic-gates for the engineering, one can create a holistic piece of arithmetical/logical machinery, which is creatively called in computer science circles the "Arithmetic / Logic Unit", or ALU. The ALU is just a

---

[6] Mathematicians and computer scientists may hate that I've given such short shrift to Boolean logic. There is much more to be said about why and how this is the case, and the equivalence/similarities between arithmetic and formal logic as axiomatic formal systems, but only the basics of Boolean logic-gates are relevant for the discussion that follows.

[7] See "long addition" and the use of "carry" columns in binary arithmetic for details on why one added to one seemingly equals zero (it doesn't). The details are surplus to demand here.
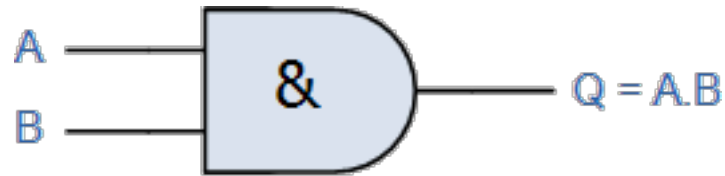
***Fig.* 2:** Graphical representation of a simplified Boolean AND-gate. **A** and **B** represent two distinct input channels leading into the AND-gate. **Q** represents the gate's output, which only takes a true value, 1, in the case that both input channels are similarly true.

	Input-A	Input-B	Output / Q
Combination 1	0	0	0
Combination 2	1	0	0
Combination 3	0	1	0
Combination 4	1	1	1

***Fig.* 3:** Truth-table for the simplified Boolean AND-gate in ***Fig.* 2**.
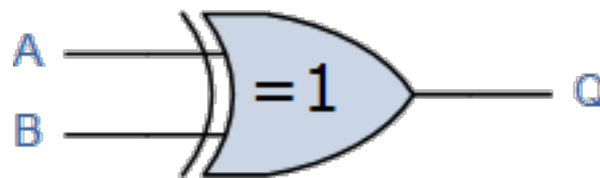

***Fig.* 4:** Graphical representation of a simplified Boolean XOR-gate. **A** and **B** again represent two distinct input channels leading into the XOR-gate. **Q** represents the gate's output, which only takes a true value, 1, when one of the input channels is true and the other is simultaneously false.

	Input-A	Input-B	Output / Q
Combination 1	0	0	0
Combination 2	1	0	1
Combination 3	0	1	1
Combination 4	1	1	0

***Fig.* 5:** Truth-table for the simplified Boolean XOR-gate in ***Fig.* 4**.
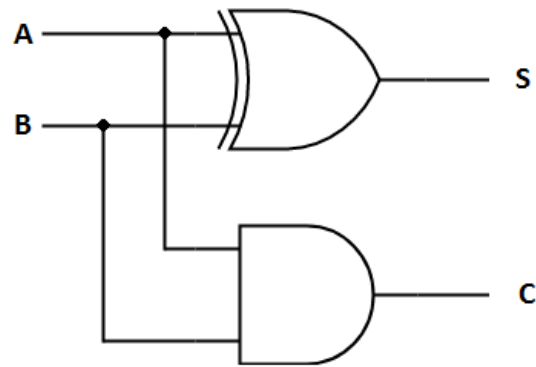
***Fig.* 6:** Graphical representation of a simplified "half-adder" circuit. **A** and **B** represent separate input channels, but note how the wiring of the circuits includes sending input to the bottom AND-gate before also reaching the top XOR-gate. In virtue of this wiring, the joint operations make the XOR-gate a "sum" channel, **S**, and the AND-gate a "carry" channel, **C**, for subsequent units to use down the line.

conglomerate of different logic-gates, like the two shown above (plus some others omitted here). The entire unit can selectively produce an output that respects an arithmetic interpretation of any given arithmetic operation one desires to perform over any given set of *N*-bit strings representing real numbers. The fundamental single-input gates scaffold together to make something more complex, like an AND- or XOR-gate, and groups of *these* gates can be further scaffolded together to make a full-adder, and *these* units can be joined to scaffold up to a complete ALU. Such is the "essence of compositionality" — from atoms to molecules, from some formal rules which govern their ordering.

To complete this story, given that digital computing machines require that the physical realisation of the symbols be separate and distinct from the physical machinery that is capable of using those symbols, there must be a *memory register* for siloing away symbols in a stable and long-term fashion. Memory registers are again simply made up of groups of logic-gates. The logic-gates that comprise memory registers are just wired together so that they can form a self-persistent value at their output rather than a merely transient one that is dependent on input. The distinct computing *core* can then withdraw the symbols when necessary in order to operate over them, using its ALU, for example. This is a general dictate of the design of digital computing machines if they are to be of any complexity and power, according to Gallistel (Campbell, 2015; Gallistel & King, 2010). The computing core can then return the results back to their relevant memory register for storage once again, or ship them off to their relevant destination for use.

Flexible and malleable *data structures* can then be built out of these memory silos. A toy-example would be a group of memory registers that store the "history of arithmetic outputs" performed by an ALU of the kind described above. Each "location" in this memory register has an *address* in binary code giving it a unique address ('0000' to '1111', for example). The first 4-bit output the ALU produces, whatever it may be, can then be *written* down in the first address, '0000', in the memory register by turning on the relevant gates in that register. As the ALU completes a new operation, the first output can be *read* and withdrawn from the first address, replaced into the second address, '0001', thus freeing up the original address. Now the new output from the ALU can be written into the first address once more. This process can iterate indefinitely as the ALU completes more operations, reading and writing 4-bit strings into the memory registers. When the memory is "full" it can simply

erase the 4-bit data stored in the sixteenth and final address as it moves all the existing data down one address, ready to put the new input into the cleared first address. This toy-example would represent a functioning data structure that represents and stores the latest sixteen outputs from a toy-ALU, as in *Fig.* 7.

The above highlights another important distinction in classical computing: the address and the data stored at that address. The address can be thought of as a 'variable', whereas the data itself represents the 'value'. In the toy-example above, the variables are categories like 'most recent output form ALU', and so on. The values within are then self-explanatory. This is incredibly important because, according to Gallistel, "The distinction between a variable (an address) and the value of that variable (the information to be found at that address) mediates the intercourse between past and present, between then and now. It is what makes it possible to bring the information gained *then* to bear on the problem one is solving *now*" (Gallistel & King, 2010: p. 154, emphasis in original). The addresses or variables stay the same, but the information within can be used or adjusted when necessary. Moreover, the addresses can become variables for effectively anything, depending on the particular system at hand and its idiosyncratic causal connections.

The above example demonstrates the core of how classical computation gets content, i.e. the values that represent and encode data regarding the variable, to be causally efficacious. The way in which the *N*-bit strings of symbols are manipulated by an ALU and memory registers respect an arithmetic interpretation. This is the first and a tried-and-trusted approach to making rigorous the concept of computation. And, while this may just appear to be computer science, rather than *brain* science, according to Gallistel, these design parameters are ineluctable constraints on cognitive science (Gallistel & King, 2010). That is, by Gallistel's lights, *if* we want cognitive science to remain a *cognitive* and *computational* domain of inquiry, we *must* locate the brain-based counterparts to the memory and processing mechanisms discussed above.

Address/Variable	Value/Information
0000	{Newest ALU output}
0001	{Last ALU output}
0011	{...}
0101	{…}

Fig. **7:** Representation of part of an idealised 4-bit memory group storing outputs from an ALU (see text above). The left column shows the unique 4-bit address for individual memory registers in the group. Each register itself can store one 4-bit string.

2.3 Rival Approaches II: Connectionism

Since the publication of the parallel distributed processing (PDP) "bibles" (Rumelhart & McClelland, 1986; McClelland & Rumelhart, 1986), there have been *two* computational games in town, despite any Fodorian rhetoric to the contrary. As discussed in Chapter 1, analog computation has of course been around for much longer, longer even than digitial computation. However, the contemporary incarnation of analog computation in artificial neural networks, and the PDP/connectionist framework that came along with it, is the most relevant to the discussion at hand regarding neurobiological computation in animals and its role in intelligent behaviour.

The classical approach to understanding information, representation, and computation, owes a heavy debt to notions and lessons learned from computer science and engineering. In contrast, connectionism's development was heavily influenced by developments in neurobiology. While there is considerable diversity among connectionist models, the building blocks always remain the same: *nodes* and *weights*. The nodes are the fundamental processing elements in connectionism, representing heavily simplified neurons. The weights are the connective tissue between the nodes, representing simplified axons and dendrites. Hence, we call groups of these nodes and their weights *artificial neural networks* (see *Fig.* 8).

The dynamics of such models are relatively simple. Each node receives input, either from external sources (the "input" to a model) or from other nodes. Depending on the type of network, nodes will usually compare a running total of this input to an internal "activation threshold". This represents a simplified version of the intraneural processes of de- and hyperpolarisation — the biological processes that dictate when a spiking-neuron will fire or be inhibited. If the activation threshold is met, the node
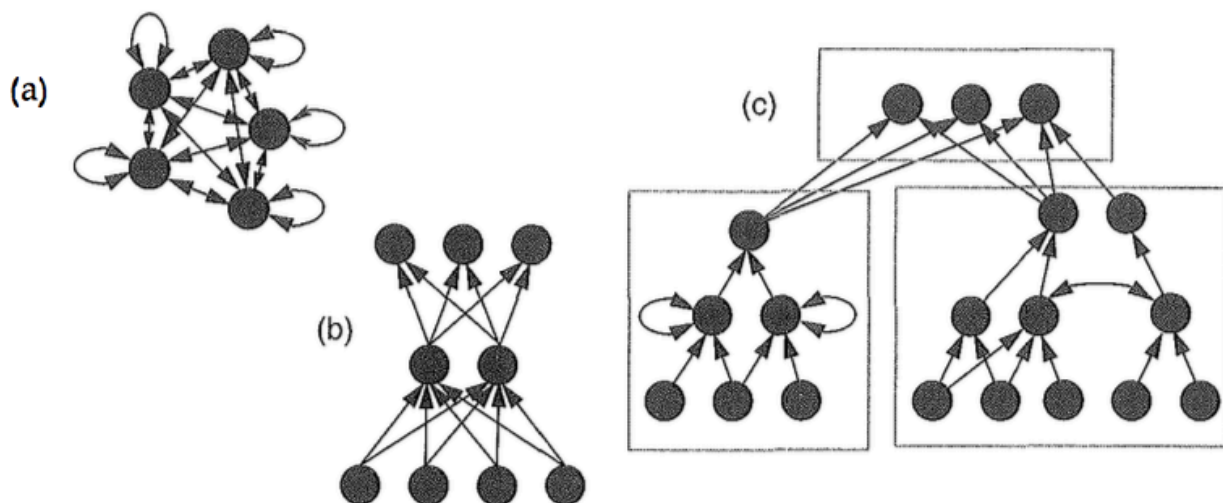

Fig. **8:** Graphical representations of various artificial neural network types. **A** shows a completely recurrent network, where the nodes' activations also feedback into their own future behaviour. **B** shows the most common variation, the three-layer feedforward network. **C** shows a complex network containing several distinct networks operating modularly together. Black circles represent nodes. The arrows indicate connections and their direction of flow of the activation or inhibition from one node to the next. (Image from Elman et. al., 1996: p. 51)

will activate and pass on its activation down through its respective connections. Other types of networks employ nodes that model graded-response neurons. These nodes react in an effectively linear monotonic fashion to their input. Rather than employing a threshold, after which activation is uniform (i.e. an all-or-nothing 'on' or 'off'), graded-responses constantly change in value to reflect the changing total input they're receiving, either up or down.

The weights act as malleable communication channels between these processing nodes. They modulate the input from one node to another by multiplying the activation value by the value of the weight. If node-A has an activation value of 1, and its connection to node-B is weighted at 0.5, the total activation received at node-B will be 0.5. If the weight value was 1, then all of the activation would be received at node-B. Weights can also extend into the negative spectrum. If the weight was -0.5, then node-B would receive an inhibitory input of -0.5. The weights thereby act as a complementary and intermediary processing phase in artificial neural networks. Weights are a simplified model of the neurobiological mechanisms that involve real synapses.

The behaviour of an artificial neural network can therefore be adjusted in two ways: change the input *to* the network, or change the weights *in* the network. The more interesting of the two is, of course, the changing of the malleable weights. A network can adjust its behaviour in future responses to the same input by altering the totality of its weight values, or the "weight matrices" between layers of nodes. The adjustment of weights between nodes represents an idealised version of the Hebbian (1949) thesis of *synaptic plasticity*: "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" (*ibid.,* p. 62). This has been affectionately repackaged as the maxim, "neurons that fire together, wire together". The weights in an artificial neural network idealise away the substantial mechanistic and metabolic changes that occur in real synapses, and merely instantiate the change in an altered weight value. An important side note is to remember that *we* model the weights using symbolic values, but they are modelling a real, physical/ structural property of a synapse. Think of the connection weight as a cylindrical channel which mitigates or permits the free-flow of activation down its length according to how open or closed it is. The symbolic value used to mark the strength of any given connection weight is therefore a representation of how hollow and open, or solid and closed, that cylinder is.

Computing with these fundamental nodes and weights takes myriad forms. Interestingly, the earliest forms also took to implementing Boolean logic, in parallel with the classical tradition. McCulloch and Pitts (1943) showed in a seminal paper how to wire together simple artificial neurons into networks that compute any finite Boolean expression. In some sense, the simple neurons with fixed weights can just be considered dressed-up logic-gates, functionally the same as the basic units in digital computing. Indeed, "McCulloch-Pitts neurons essentially *just are* the logic gates in conventional computers" (Morgan, 2014: p. 84, emphasis in original).

However, the important developments came when the introduction of intermediate layers of more complex nodes, with nonlinear activation functions and rules for adjusting malleable weights, began in earnest. These changes in complexity added temporally dynamic behaviours to the networks, but the most important change was one in interpretive logic. Rather than considering the nodes as atomic and
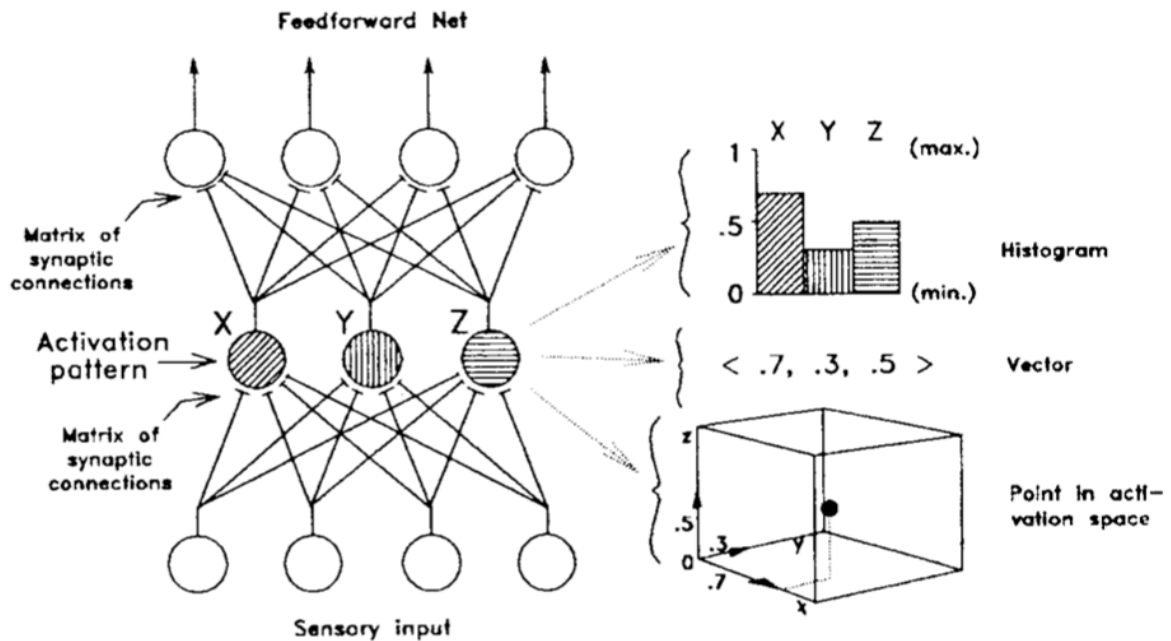
Fig. 9: Three ways of visualising the concurrent activity of a population of artificial nodes. The total activation at the middle layer (left) can be conceived as a histogram or pattern of activation levels (top right), an activation vector (middle right), or as a point in activation space. (Image from Churchland, 1998: p. 6)

neurobiologically inspired logic-gates outputting symbols, the *distributed processing* that occurs in *parallel* across all nodes in a given layer of a network became the focal point. Consider ***Fig. 9***.

The primary vehicle of representation and computation under consideration here is the overall *pattern of activation* at the intermediating layer of three nodes (*X, Y*, and *Z* in ***Fig. 9***). Each node will give some value of activation, varying usually from 0 to +1. Visualising this activity as an undulating histogram of values makes it easier to see it as a collective "pattern" of activity. Alternatively, they can be characterised as an activation "vector": an *n*-tuple of values where *n* is equal to the number of nodes in the layer.

Adopting the mathematical language of vectors allows a further interesting development: the creation of an abstract geometry with which to analyse these patterns of activation. If we take each node as a dimension, an axis across which values can vary, a space can be created within which all possible combinations of values can exist: the *activation space* (see bottom-right, ***Fig. 9***). Each unique combination of activation values, each pattern, can then be plotted to its accompanying point in this space. The values of the nodes become coordinates in space under this description, "physically described" rather than formally described as some *n*-tuple (Churchland, 1998).

This way of analysing patterns of activation is special as it allows a *spatial* metric of analysis: *distances between locations* in activation space. Each point in activation space acquires its computational salience in virtue of its place *relative* to all other contentful points in that same space, *not* merely by its position relative to the constitutive dimensions of that space (Churchland, 1998). Consider Churchland's (*ibid.,* p. 9) example of two identical, but imagined three-layer feedforward networks that discriminate faces from a set of photos of faces (***Fig. 10***).
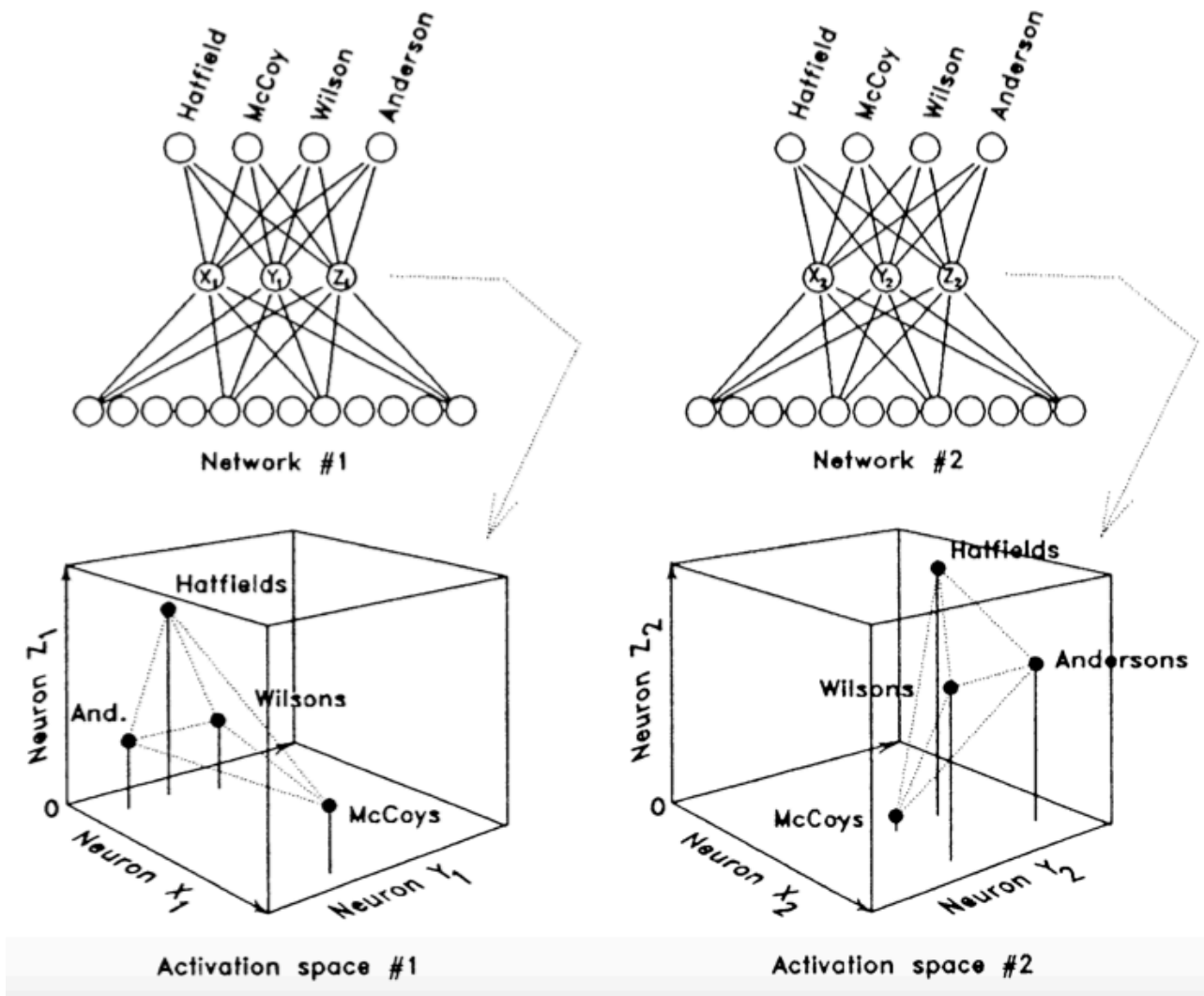
***Fig.* 10:** The activation spaces of two identical imaginary networks trained to recognise faces. The networks are trained to identify any given input, a photo of an individual face, and output which family of faces it belongs to. The points marked in the activation spaces represent "prototype points", around which any given photo of a single family member will cluster. Note the same essential geometric structure that is created by and between the four prototype points. See text for details. (Image from Churchland, 1998: p. 9)

The networks are given as input one photo out of one-hundred possible photos of faces. Each face belongs to one unique member of one of four families. In total, there are twenty-five Hatfields, twenty-five McCoys, twenty-five Wilsons, and twenty-five Andersons in the input set. Some details regarding training aside (see Churchland, 1998: p. 9), by altering the weights in the networks they can be trained to recognise each photo and output which of the four families the member belongs to. The output is defined as the highest of the activation levels across the four output nodes.

One can easily imagine that faces can be broken down into several dimensions, like nose length, eye size, jaw width, and so on. Given that families share heritable phenotypical traits, any given Hatfield member will, across the facial dimensions chosen, score closer on average to the other Hatfields than a McCoy or a Wilson. If we averaged all the Hatfield's scores, for example, we could create a non-existent "prototype Hatfield", and this would differ from the averaged prototypes of the other families. When actual artificial neural networks are trained to perform such tasks as the above, they produce exactly the kinds of results as displayed in *Fig.* 10.

To reach the desired performance at the task, networks are given small "nudges" across trials, fractional adjustments of the initially random set-up of connection weights between nodes. This a kind of "supervised learning", and each iteration of adjustments gets them closer to the goal. This process carves up the structure of the network into distinct spaces. Once the network reaches some desired accuracy level when completing the task, the activity histograms of the middle "hidden" layer of nodes in response to each of the 25 family members can be plotted into an activation space. Furthermore, these responses could then be averaged. The resulting averaged coordinates create exactly the kind of "prototype" of one family's faces discussed above. Performing this analysis on all the families reveals how the network's activation space is partitioned into familial zones.

The two imagined networks ended up with different encoding strategies in solving their common problem, due to their differences in random initial set-up. This is demonstrable by the unique positions the prototype points take in each network (see bottom *Fig.* 10). However, the *relative* positions of the prototype points—with respect to one another—are geometrically similar. Indeed, in this toy-example they're identical. The points form an irregular tetrahedron in *both* networks, just rotated and translated relative to each other.

This demonstrates the point about *relative* distances between locations in activation space as the salient principle in artificial neural network computation. The relative differences and similarities in facial dimensions, between and across family members, is being mirrored in the abstract structure of the networks. The latter is a physical *analog* of the former. The process of nudging the networks connective tissue forces the bottom section of the networks to convert those facial dimensions into a much leaner amount of information across just three dimensions, the middle three nodes. Each middle node takes a variable amount of modulated activity from the input nodes, according to the final set-up of the connection weights. The top half of the networks then decode that condensed information into the relevant familial pigeon holes, exploiting the difference/similarity relations created within the bottom section. No individual node carries "information" by itself, barring the arbitrarily defined output nodes (for simplicity in modelling). The collective activity between input and output distributes that load instead, in parallel. Each image of a face is encoded into a triplet of activity, and the

collective group of triplets form a relational structure whose geometry is sufficiently similar to (at least some of) the relational structure of the real faces under examination. Like the orrery from Chapter 1, by leveraging similarity and difference relations in their structure, the networks thereby form and exploit a *structural resemblance relation* with the task domain.

In sum, we can identify a prototype point in a network by its causal connection to the input, e.g. we can just observe the standard averaged response to any particular family. This is easy enough. However, "What matters is not [a group of prototype points'] accidental location or orientation within any given space; what matters is its internal geometry, and the causal connections of its various vertices to objective features of the sensory environment" (Churchland, 1998: p. 12). This is why we call such artificial neural network models *connectionist* models. The interpretive logic is a shift away from classical thinking, regarding symbols and logical functions, to distributed patterns and connective structures.

This is the *second* game in town, computationally speaking. Over the last few decades, it has also become a tried-and-trusted approach to completing myriad computational tasks. The next section will turn to analysing how these different approaches to computational hardware can be said to be deploying "representations" in more detail.

2.4 Symbols, Structures, & Homomorphisms: Philosophical Problems

This subsection will analyse the different content determining relations that are required to separate the competing computational theories described above. It will be a story of two different kinds of *homomorphism*: functional homomorphism and structural homomorphism, respectively.

Gallistel (1998) writes:

> A mental representation is a functioning isomorphism between a set of processes in the brain and a behaviorally important aspect of the world. This way of defining a representation is taken directly from the mathematical definition of a representation. To establish a representation in mathematics is to establish an isomorphism (formal correspondence) between two systems of mathematical investigation (for example, between geometry and algebra) that permits one to use one system to establish truths about the other (as in analytic geometry, where algebraic methods are used to prove geometric theorems).

Homo- and isomorphisms are, as Gallistel highlights, a notion recruited from mathematics. These are sometimes used interchangeably, depending on how strict an author wants to be, and *iso*morphism is the stricter of the two relations. I take an ambivalent stance towards the difference here, and will use them interchangeably. The core concept is that '*homo*' and '*iso*' both mean 'same', and '*morph*' means 'form' or 'structure'. So, a homo- or isomorphism is a structure preserving mapping from one structure to another.

An easy example to visualise a "structure preserving mapping" is between geometry and algebra, as Gallistel suggests. The Pythagorean theorem—that A^2 plus B^2 is equal to C^2—is perhaps one of the easiest examples to picture. The formal theorem says true things about the geometric domain. By manipulating algebraic structures, we can thus reason about geometric structures, because the way that

A^2, B^2, and C^2 relate to each other sequentially in the theorem (their formal structure), and then causally relate to each other during the manipulation process (their causal structure), preserves something about the way that the real sides in a right-triangle relate to each other physically.

As discussed in sections **1.2** and **1.3,** using one system of elements to examine and come into contact with another domain is the core of representationalism. However, Gallistel emphasises a *functioning* homomorphism, between "a set of processes" and some relevant aspect of a task domain, as the underlying principle that permits a representational interpretation of such processes and their constitutive elements. This requires that:

i)   there is a homomorphism between *A* and *B*.

ii) the homomorphism between *A* and *B* is sustained by causal relations between them, such that variations in *A* are sensitive to and mirror the variations in *B*.

iii) A is exploitable by some user-system to guide its behaviour with respect to B, in a way that reflects the relevance of B to the user's behavioural milieu.

We can map the variables in Pythagoras' theorem to the relevant sides of a real right-triangle, so i) is satisfied. Condition ii) is satisfied by virtue of the rules of arithmetic, i.e. the order and type of causal manipulation performed on the values of the variables. We can put the rules of arithmetic into some computational hardware, e.g. using a combination of an ALU plus memory registers as shown in **2.2**. As a result, the transformations of the values in that hardware can guarantee some user-system an output that can be interpreted as respecting the geometry being computed over. Condition iii) is therefore satisfied too. One can use the theorem to solve for the length of a hypotenuse of a right-triangle because these three conditions are satisfied in the homomorphism between algebra and geometry.

It becomes clear from this picture that what Gallistel really means by a 'functioning homomorphism' is a functio*nal* homomorphism: a structure preserving mapping established and sustained by the *causal function* of the elements within the representing system. This is not surprising, as this is how symbols get their content. Symbols represent only in virtue of their functional role, because symbols qua physical tokens only arbitrarily relate to their represented objects. However, symbols can be manipulated by formal operations in a holistic *system* of symbols, whereby the *causal liaisons* of those symbols can be made to be homomorphic with certain patterns of causal liaisons present in the task domain. For example, it's not that any one symbolic encoding of a value intrinsically stands in for one individual side of a right-triangle. Rather, the *system* of symbolic encodings can be made to respect a trigonometric interpretation if their causal relations are disciplined in a certain way. The way that the values encoding A^2 and B^2 causally function within the system preserves the formal relations between the geometric structures that they're mapped to. This is par for the course in classical computer science, and by Gallistel's lights, so it is for cognitive science: representing systems are symbolic systems, necessarily, as only symbol systems can sustain functional homomorphisms — and only functional homomorphisms count as genuine content determining relations.

However, this is not the only way to reason about geometry. There are physical magnitudes that can be manipulated in order to perform the same work. That is, *geometry* can be used to reason about geometry. Take a right-triangle you're faced with, duplicate it three more times, and arrange these so that the hypotenuse from each triangle constitutes one side of a square (***Fig.*** **11**). One can see what the process of "squaring" literally means: it's the area of the square formed by using the length of that side. Move the triangles together so that you are left with two squares of empty space now, formed individually by the lengths of the sides you know the values for (see bottom ***Fig.*** **11**). Given that the overall space hasn't changed, one can reason that this area, A^2 plus B^2, is equivalent in area to C^2. Thus, one has reasoned about geometry, i.e. a line (the hypotenuse), by manipulating other geometrical structures.

This may seem a bit trivial, using geometry to reason about geometry. Consider though what sustains the usefulness of the shapes created and manipulated in ***Fig.*** **11**: it is the real and intrinsic physical properties of the shapes, *not* their function. What fixes the content of the representations of geometric features in this example are their own intrinsic geometrical features.

So, remember again that our toy networks did not use elements that *function* homomorphically to the elements in their task domain. The networks successfully pigeon-holed faces by having their underlying connective structure, and thereby their activation patterns, partitioned in such a way that their intrinsic abstract *geometrical structure* resembled some relevant counterpart structure in the task domain. In the above conditions for a homomorphic relation, i) and iii) remain the same. However, the condition in ii) that the homomorphism be sustained by *causal relations* can be substituted with the condition that:

> ii*) the homomorphism between *A* and *B* is sustained by *structural relations* between them, such that variations in *A* are sensitive to and mirror the variations in *B*.

The structural relations that constitute the irregular tetrahedrons in the toy-networks' activation spaces are sensitive to and mirror the variations in structural relations between their input. This kind of 'structure' is a real physical structure created by the differences and similarities of the activation patterns, analysed synchronically as their positions and distance relations in activation space. This is a kind of *second-order resemblance*: the physical relations in the representing system resemble the physical relations in the represented system. The activation patterns don't individually resemble the faces' first-order properties. However, the *relative* differences and similarities of the activation patterns resemble the *relative* differences and similarities of faces and families.

A homomorphism can thus still be mapped between each activation pattern and each individual face from the input. However, the meaningfulness of this homomorphism is not sustained by the causal or functional relations between the activation patterns, as there are none to speak of. The homomorphism is sustained by the physical geometry of the irregular tetrahedrons in the networks' activation spaces: the distances between points, relative to one another. Leveraging second-order resemblance relations, i.e. physical analogies in a structural homomorphism, is the underlying principle of 'structural representation' (see O'Brien & Opie, 2004; 2005).
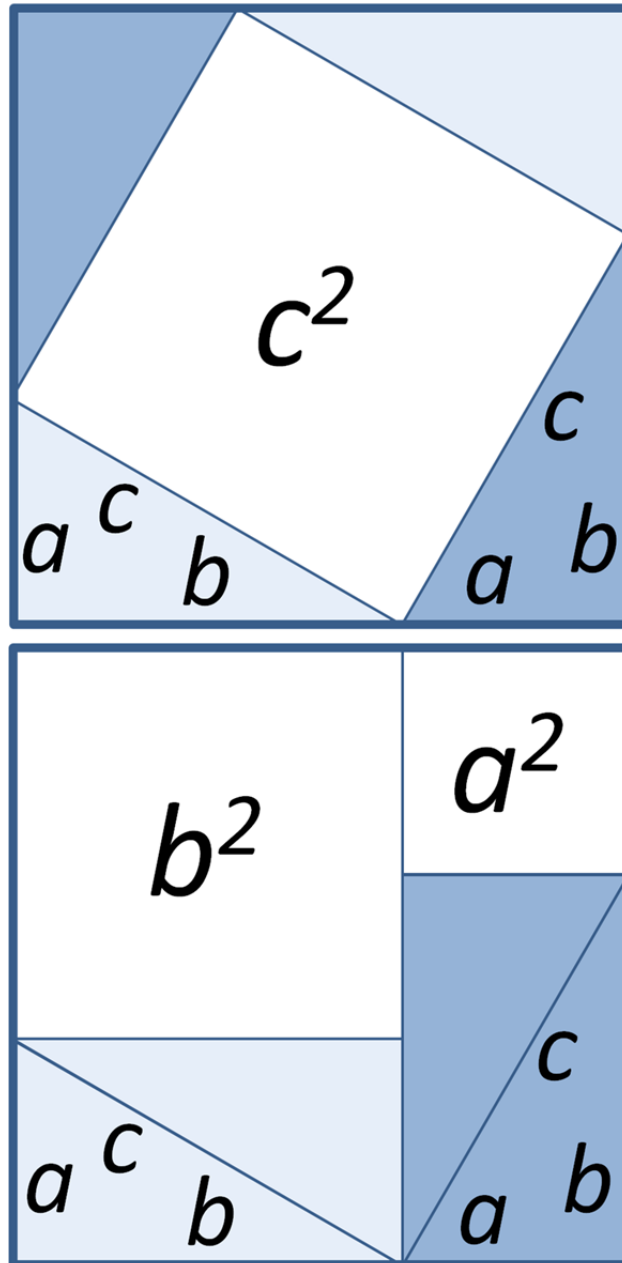
Fig. 11: One geometric proof of the Pythagorean theorem. This example demonstrates how the intrinsic geometry of shapes can be manipulated and exploited to reason about certain tasks. In this case, one can see that reproducing a triangle three more times can allow one to visualise the hypotenuse "squared", literally. Subsequently, moving the triangles, top-left down to bottom-right and top-right down to bottom-left, forms different shapes and areas that allows one to reason about the relationship between the lengths of the sides to the hypotenuse. See text for details.

This gives us a picture of connectionism as a non-symbolic mode of computation. Instead, the nodes in a network, their activity patterns and their weighted connective tissue, are understood as sustaining a structural homomorphism with their task domain. Connectionist networks leverage these structural representations because their causal/computational elements are one and the same with the elements that sustain the structural homomorphism.

Connectionist networks are thus analog computational systems. In analog computers, "…it is the existence of a relation-preserving mapping between *physical properties* of the representing vehicles and the modelled system that grounds the content of those vehicles" (O'Brien & Opie, 2008: p. 9, emphasis added). We have now seen several examples of what a relation preserving mapping sustained in physical properties would look like, from orreries to face-recognising networks to moving geometrical shapes. Connectionist networks are therefore using geometry to reason about other geometries, and formal algebraic reasoning gives way to the exploitation of physical analogies that exist between the relational characteristics of physical structures. This is an altogether different approach to representation and computation from the digital, classical, symbolic approach.

In sum, symbols must be made to *behave* in a homomorphic fashion to their task domains in order to be reasonably interpreted as representing that domain. Structural representations, on the other hand, exploit an intrinsic structural homomorphism that exists between their own relational/geometrical structures, and that of their task domains. A functional homomorphism is a different relation to a structural homomorphism, but either could exist between elements of a representing set and the elements of a represented set. Gallistel has clearly shoe-horned one kind of homomorphism into his general conception of homomorphisms, and how they relate to representation. This is unsurprising given that he favours the classical, digital, symbolic approach to computation. If you shoe-horn symbolic computation into your general notion of computation, you consequently arrive at functional homomorphisms in the same way: at the expense of other approaches. However, we have seen that there is a separate approach to homomorphism that can be at least as robust a representational relation as the other, and how it can be leveraged in a computational system. The difference in which homomorphic relation is being appealed to, and thus which kind of representation is being invoked, is *the* difference between classical and connectionist computation.

2.5 Computational Tasks & Logical Framing: Engineering Problems

This subsection will sketch some engineering consequences that result from applying the different representing media illustrated above. Given that they require different content determining relations, they will depend on different properties of their underlying substrate: either structural or functional properties. As a result of this, the two approaches will, at least some of the time, require different "framings" of the same tasks, to properly leverage their respective representational media towards completing that task.

This section's discussion is about *task dependence*: how you go about completing a task depends, in some measure, on the intricacies of the task itself. For example, we saw how reasoning about the relations between sides in a triangle using different representing media resulted in different approaches to operating over those media. One approach used arithmetic in solving an algebraic formalisation, the

other manipulated physical structures in space. However, given the task itself, not just any physical structures would fulfil that task. We need to put this notion more concretely into a biological and behavioural context though.

There is universal acceptance that extracting information from the environment informs our later behaviours. However, any given aspect of behaviour is informed by elements of past experience that may have happened at very different times, and may have been collected from different sources. All of these encoded pieces of information must be brought together at the relevant moment for the relevant task. What kind of information you collect will trivially constrain what kind of work you'll be able to do with that information. That is, which aspects of your environment your sensory apparatus is actually sensitive to will constrain your logical framing of the problem at hand, because it constrains what kind of structures you have available to play with.

Take the problem of navigation. This is a ubiquitous problem for moving animals who venture from home to search for and retrieve food, for example. Within this task is the sub-task of *location*. There are many ways to represent one's own location in space. Humans solve this problem of pinpointing location using one of two common systems: 'Cartesian' or 'polar' coordinates (see ***Fig. 12***).

Cartesian coordinates define a point in space by its distance across some fixed perpendicular dimensions. Take two-dimensional movement for now, just across the *X* and *Y* axes. The reference point, or origin, is represented as (0, 0), the point at which the two axes meet. If I started there, and moved five units across, and ten units up, my location is now the point defined by (5, 10) in the Cartesian plane. These units could be anything; metres, centimetres, or natural strides even, just so long as they're both the same.

On the other hand, polar coordinates define a point in space by its distance, ***r***, from a reference point (the *pole*), and its angle, ***Θ***, away from a reference angle. Polar coordinates, therefore, define a *line* between the reference point and your location. These are different representations of the same problem,
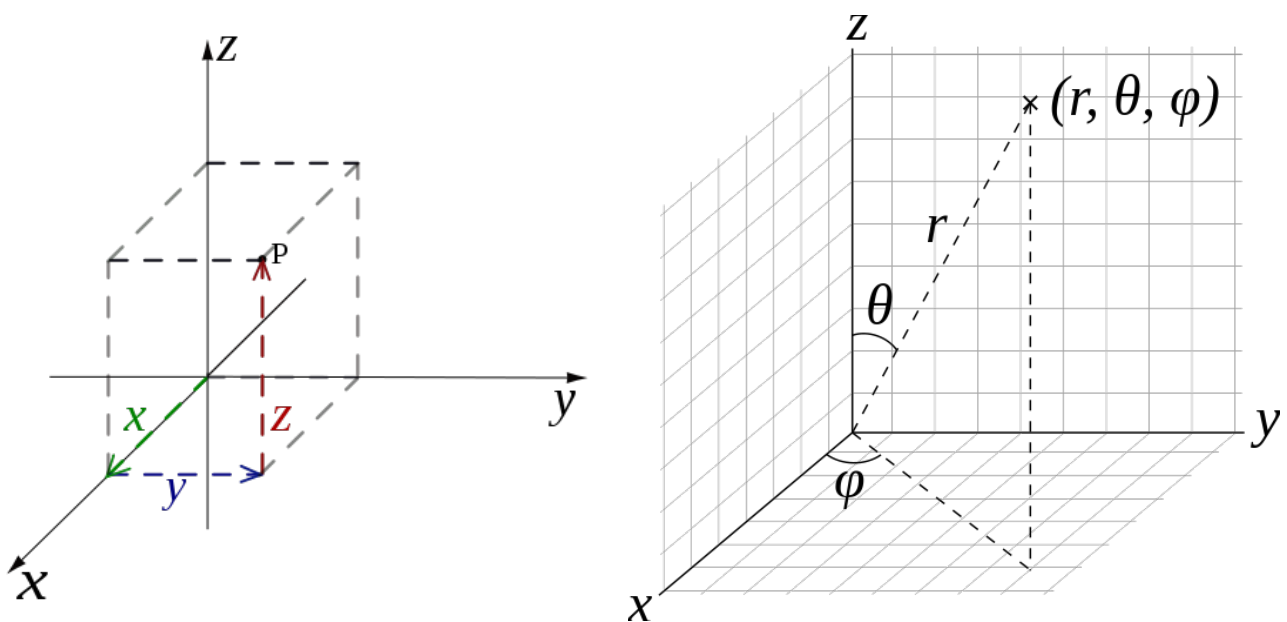

Fig. 12: Graphic representation of two coordinate systems, Cartesian (left) and polar (right). Both can be used to represent the location of a point in space, relative to some axes that define the dimensions of that space. Details in text.

but the same kinds of information need to be collected to solve it. However, different types of data structures can be built depending on *how* that information is represented.

Both coordinate systems require a sensitivity to orientation, and a sensitivity to magnitudes for distances. In the Cartesian case, I must keep track of which axis I am travelling in, and how many units I'm travelling in that axis, in order to update my coordinates (*x, y*). For example, if I move diagonally north-east for one unit, this can be represented as (1, 1) in the cartesian plane. Moving diagonally north-west for one unit would be (-1, 1). I can then keep a running two-column tally to update my coordinates. If I'm moving across the east-west axis, I add or subtract units to the *x* column, and if I'm moving across the north-south axis, I add or subtract units to the *y* column. This method exploits the functional homomorphism between algebra and arithmetic with the underlying geometry I want to compute over. Take an ALU and memory registers from **2.2** (see *Fig.* **7**), and one can build a flexible symbolic data structure to track and compute the changes in my location using simple arithmetic.

For polar coordinates, I similarly need to track a length of travel, and the angle of that travel, to update my coordinates (*r, Θ*). However, consider if I instead used lines on a grid to represent these pieces of information. Rather than taking in data regarding my strides, and representing it symbolically, I can represent it using another physical magnitude: the length of a line drawn. Similarly, I can take in data regarding the angle at which I was travelling, and rather than representing that as a symbolic value (in degrees or radians), the angle at which the line is drawn — outwardly from the pole and relative to all other possible angles — signifies this information. Iterate this process for each path I walk, and what accumulates are groups of lines on the grid. This method exploits a structural homomorphism between the geometry of lines on a grid, and the paths taken in actual space. Note that, in contrast to the Cartesian case, this is a non-symbolic, non-arithmetical way of computing my location(s) in space. It is a different system altogether, and is manipulated in a different manner.

How exactly I represent values for these variables of distance and orientation thus ultimately determine what kind of representing system is better to build and operate over. Perhaps I can only count my discrete strides, in which case I may turn to using a symbolic data structure as in a two-column Cartesian coordinate system. On the other hand, perhaps I use a stop-watch to time each section of travel, in which case I may exploit a polar coordinate system to build lines on a map, such that the lines' lengths are proportionate to some set amount of time. For orientation, if I have a compass that shows me where in the cardinal directions I am facing, I may build a Cartesian system. In contrast, perhaps I'm in an unfamiliar place, but there are distinct features in one particular direction that can be used as landmarks. I may leverage this feature to then estimate my directions of travel calibrated as angles of deviation away from the direction of those landmarks, as in a polar coordinate system.

All of these considerations have non-trivial effects on how we logically frame, interpret, and build systems to solve problems. This is the case whether they are automated systems, or systems we build with pen and paper. The kinds of variables required to compute over can take various representational forms. Those representational forms affect the kinds of systems we build, and those different systems come with different methods of manipulation. The rest of the thesis will now analyse one such biological example where all of these philosophical commitments and their engineering consequences dovetail, in a striking case of animal navigation.

Chapter 3

Path Integration & Polarisation-Sensitive Vision in the Desert Ant

A Case Study in Basic Cognitive Phenomena

"Ants are so much like human beings as to be an embarrassment. They farm fungi, raise aphids as livestock, launch armies into war, use chemical sprays to alarm and confuse enemies, capture slaves, engage in child labour, and exchange information ceaselessly. They do everything but watch television."

- Lewis Thomas

3.1 The Evolutionary Niche of the Desert Ant

The intriguing foraging behaviour of Desert Ants, in particular the various species of the genus *Cataglyphis,* has been used by C. Randy Gallistel as a paradigmatic example of simple cognitive phenomena in animals. Gallistel has gone so far as to say that Desert Ant foraging behaviour is, "*…the* central behavioural phenomenon that people should focus on if they're interested in the neurobiology of computation" (Campbell, 2015). It is intriguing because it is "simple", in that the ants possess very few neurons, especially when compared to humans — the cognitive animal *par excellence*. Nonetheless, their behaviour seems to require a sophisticated *computational* explanation. We can understand why it is so sophisticated by first looking at how the evolutionary niche these ants inhabit, i.e. the Saharan Desert floor, imposes several unforgiving constraints on their behaviour. Looking at these factors will motivate and make clear why theorists, like Gallistel, believe Desert Ant foraging behaviour is an intriguing case study to test the explanatory power of our theories of cognition against the world.

The most important of these environmental conditions is the remarkable temperature range of the desert. The extreme heat, by a confluence of factors, ultimately sets a time constraint on the ant's behaviour. Firstly, the surface temperatures on the desert floor reach up to 70°c. This forces the Desert Ant to be a carnivore, because not a lot of plant-life can grow in such conditions. It must therefore leave the proximity and security of its nest, braving the extreme temperatures, as a solitary hunter searching for other dead animals for food. Typically, they eat other arthropods that have died while attempting a similar venture. However, the absence of plant material naturally also makes other species who share the habitat carnivorous. The Desert Ant must be able to avoid such predation.

To avoid predation from any larger carnivorous creatures, *Cataglyphis bombycina*, for example, only ventures from its underground nest in search of carrion in a short time frame around noon. At this time of day, when the temperature is highest, all the other predatory species retreat to avoid the blistering sun. The desert floor is thus only clear for the ant to hunt in solitude and relative safety around noon, exactly when it is at its most inhospitable temperatures. The ant's behavioural adaptation[8] for surviving these extreme temperatures is to restrict its search time — outward and homeward inclusive — to around ten minutes. If the trip lasts much longer, the ant inevitably dies of heat exhaustion, just like its intended food source.

The upshot of all of these heat related constraints is this: the ant needs a *fast* method of navigating, one that provides adequate search time, but also minimises the homeward trajectory. It needs a method that allows it to maximise its chances of finding food without incurring extra risk, where an additional unit of outward search time does not necessarily add a unit of homeward travel time.

Two other constraints limit the possible types of fast navigation available to the ant. The desert plains are windy, which makes the dropping and retracing of sensory trails useless; and the desert floor is

---

[8] Many genera of ants have developed various other biological adaptations that aid in surviving the heat. These include protective coats to disperse or reflect sun radiation, and specialised proteins that support cellular function above the normal critical thermal maximum for short periods of time. Of course, only the behavioural adaptations are important for our analysis of their cognitive capacities.

effectively featureless, which makes any form of long-range landmark navigation unviable. All that is left to navigate with is the most predominant feature in the environment; the sun.

3.2 Path Integration & Celestial Navigation

If Desert Ants don't use "landmark navigation", they must use another technique, called "path integration" or "dead reckoning". The former strategy is an example of *allothetic* navigation, the use of relationships between external cues, whereas the latter is *idiothetic* navigation, the use of cues that are internally generated (Whishaw & Brooks, 1999; Wehner & Srinivasan, 2003). These require distinct computational strategies. The goal is the same, i.e. to integrate one's final position and then reach home. However, the information that is available to be processed is radically different.

Path integration is a technique familiar to human seafarers. This is because the open ocean is analogous to the desert floor: it is a featureless and homogeneous expanse, and the movements of the water are analogous to the effect that the wind in the desert has in preventing the use of trails as information cues. "Path integration" is the use of some effectively stationary features of the sky to overcome this poverty of terrestrial (or nautical) information cues.

Celestial bodies are essentially at infinite distances for observers on Earth. This means that any translational movement by a terrestrial navigator —movement in the standard three dimensions— will not affect their observed positions in the sky. Only rotational movement —spinning around a centre point which does not change — will affect their observed positions. Of course, some of these navigationally useful celestial bodies, e.g. the sun, trace variable trajectories across the sky over the course of a day and year. This means that any sufficiently long journey that uses the sun's position in the sky must take into account that positional change. So, by matching for specific times, one can use these regular and/or fixed positions/trajectories of celestial bodies to determine one's direction of travel at any time.

This is making use of *celestial ephemerides*; functions that describe the positions which naturally occurring celestial bodies will take in the sky at particular times of day and/or year. Historically, these were printed and tabulated into almanacs, a look-up table, in essence. This highlights the etymology of the term; 'ephemeris' is both Latin and Greek for 'diary' or 'journal'. After keeping track of how far (distance) or how long (time) you travel in particular directions using the celestial bodies as compass markers, you can sum this information to compute your net displacement relative to point of departure. Travelling home is now fast and easy. Rather than retracing all of the parts of your outbound journey, which could be numerous and lengthy, just travel in as straight a line as possible from your final destination point to home.

This kind of navigation is exactly what is observed in Desert Ants (see *Fig.* 13). Upon finding a source of food during a foraging trip, the ant will take a more or less straight-line path back to its nest, regardless of the complex or seemingly random search trajectory it has taken. Given that it doesn't retrace its outbound path, evidently it doesn't follow any sensory trails left on the way out, avoiding the problem of wind. However, neither is it using proximate environmental features as periodic cues with which to pilot its way home. The Saharan Desert floor is, after all, barren of any suitable such features. So what information are they using to perform such a sophisticated computation?
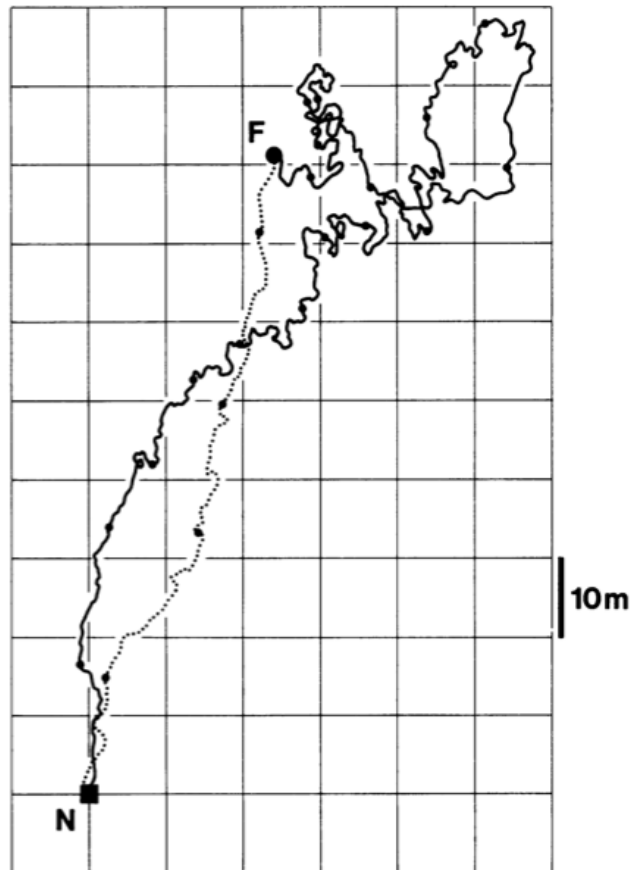
Fig. **13:** Graphical representation of the foraging journey of an individual ant, *Cataglyphis fortis*. The foraging path is represented by the solid black line, the return path is represented by the dotted line. *N,* nest; *F*, feeder location where the ant found food, marking the beginning of the return path. Small solid circles represent time marks, given every 60 seconds. (Figure from Müller & Wehner, 1988: p. 5288)

Experiment demonstrates that Desert Ants will systematically *mis*represent the location of their nest if they are displaced by experimenters at the moment they find a source of food (Wehner & Wehner, 1986). Moving them will cause the ant to miss their target by a distance and angle that corresponds to the distance and angle of displacement. For example, if their final destination was north east of their nest, and they are displaced to the south west of their nest, they will continue to walk south west, farther away from their intended destination, instead of correcting their path by heading north east. This is a very systematic and proportionate way of committing errors.

They will also stop and visibly demonstrate "confusion" at having not found their nest by changing their straight-line pathing behaviour back into a systematic search procedure. This indicates that they hold an internal representation of a homeward *path*, i.e. something that has a finite *length* as well as direction, and are aware of when that path has been travelled (Wehner & Wehner, 1986). If they were using external cues, one would expect the ant simply to recognise how their final position had been altered by the experimenters, and adjust for this accordingly by piloting for as long as necessary. Instead, they simply walk to where their nest *should* have been, had they not been interfered with.

On their very first foraging trip, the terrain they cross on the path home could be entirely novel territory that wasn't covered on their outbound trip, further reinforcing the notion that they are not using chemical signposts or landmarks to reflexively navigate with. This instead favours the hypothesis that they use some form of path integration and dead reckoning, at least part of the time.

We now have some initial broad strokes of what informational states the ants are exploiting. The proportional way they misrepresent in the experimental conditions described, combined with their distinct "halting" of path-making behaviour, strongly suggests that they are internally representing and computing over *vectors*[9] specifically, not any other absolute and external spatial relations. The ants are, therefore, performing some biological approximation of vector addition[10]. Vectors are quantities that possess both direction and magnitude information. We commonly represent vectors graphically as an arrow, as arrows concisely capture the notions of directionality and length. For example, 'velocity' is a vector, in that it represents the *rate* and *direction* of displacement, i.e. how fast an object is travelling, and in what direction.

For navigation purposes, there is an obvious and natural relationship between one's velocity and one's path-making, because the latter can be derived from information about the former. More time spent travelling in a particular direction overall, or travelling at a greater speed, entails a greater distance covered in that direction. That is, it leaves a longer path laid down in one's wake. Straight-line paths, once laid down, are also describable as vectors: they begin at a point *A* and persist to some point *B*, giving them a natural 'directedness' and length. At some suitable scale, each small subsection of the outbound journey of the Desert Ant (see *Fig*. 14) can be described (or approximated) as a short linear path. The overall computational task involved in computing net displacement is therefore one well suited to the mathematical language of vectors and vector addition.

A wealth of empirical evidence supports the model of Desert Ant navigation by path integration and vector addition. Most of this data is informative because of the systematic way that animals who exploit this navigational method make navigational errors. These systematic errors are manipulated and observed when the creatures are artificially exposed to redacted, or otherwise suboptimal information that would normally not occur in the regular and natural conditions of their foraging behaviours.

For example, the method that ants use to gauge and keep track of distance (the magnitude component to path integration) was experimentally tested by artificially lengthening or shortening their legs as they were beginning their homeward journeys (Wittlinger, Wehner & Wolf, 2006). When the legs of a foraging ant are surgically shortened they systematically undershoot their target. The opposite result is found when their legs are lengthened. This suggests that proprioceptive signals called 'efferent copies', that fire whenever movement of the legs is initiated are being used as an idiothetic measure of distance. This makes use of the natural relationship between velocity and path-making; taking longer strides or more frequently engaging movement of the legs entails a greater distance travelled. Shorter legs require more time to travel the same distance than longer legs, assuming the same speed of engagement. Efferent signals are thus an accurate and accessible proxy for measuring distance, at least in natural conditions where leg-length remains constant. The use of efferent signals in ants has been

---

[9] Throughout this discussion I use the term *vector* in this sense of a *physical vector*, or the mathematical quantities used to describe them. Connectionist literature, as in **2.3**, sometimes refers to activation patterns across a given layer of artificial nodes as 'activation vectors', and this is not to be confused with the sense used here. While these uses do indeed share some similarities, to avoid confusion I strictly refer to vectors from now on as only those quantities that possess direction and magnitude in the *physical* sense.

[10] Path integration or dead reckoning is also referred to as 'vector navigation' in some corners, which reflects the underlying mathematics involved.
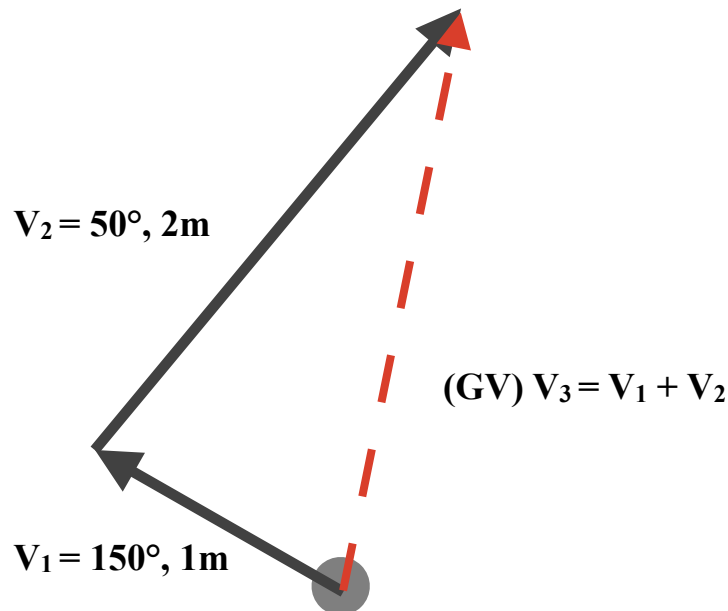
Fig. 14: A simplified example of path integration, showing the relevance of mathematical concepts such as vectors, and trigonometry. If a creature were to traverse an idealised path describable as following the vectors V_1 and V_2, it could derive its net displacement by summing these two quantities together to form a third, 'global' vector, represented by the broken line V_3. Inverting the global vector represents the straight path required to get home. The nest, or point of departure, is represented by the grey circle.

affectionately (but very loosely) referred to as "step counting". A more accurate metaphor would perhaps describe it as the laying down of a thread, rather than the numerically loaded term "counting" (see Wehner & Wehner, 1990). Wittlinger et. al. (2006) dubbed the precise mechanism as the "stride integration hypothesis".

The method of gauging and tracking direction information in ants has been illuminated by two similar types of experimental designs. Both introduce a kind of suboptimal or "poverty of the stimulus" condition, where elements of the relevant stimuli are redacted. One type removes temporal information, the other removes spatial information. The way the creatures reliably react to the systematic variation in these conditions sheds more light on their representational apparatus, and therefore, the specific computational strategy they employ. These experiments, and the relevant underlying neurobiology involved, however, are much more complicated than the method for gauging distance, both in a biological and computational sense. But they are are vital to the discussion. So, the next sections will begin looking at these finer empirical details, in order to get a better grip on just how the ants successfully navigate their difficult environment via path integration and vector addition.

Understanding the evolutionary context of the Desert Ant, and the cursory behavioural data above, is thus useful in narrowing the problem domain down to a certain extent. At the broadest level of analysis[11], this confluence led us to the conclusion that path integration must be their overarching navigational method. Similarly, a more detailed understanding of the neurobiology involved in the determination of directions of travel using celestial information will help us narrow the problem down even further: what specifically are they representing about their paths, directions, or locations? These

---

[11] I use Marr's (1982) levels as a heuristic for what I mean by "levels of analysis". I don't specifically endorse Marr's account, however.

are the practical questions that sit parallel to the ultimate points of contact between connectionism and classicism, and between my position and Gallistel's: how does something neurobiological instantiate something representational? However, this will require a somewhat detailed understanding of the features in the environment available for the ant to compute with in the first place, i.e. the way in which the sun and sky relate to abstract variables like 'direction' or 'time'. Only then can the discussion move on in subsequent chapters to modelling how such a celestial compass system can be integrated with idiothetic signals for distance, and ultimately exploited to compute a global home-vector. That is where the rubber really meets the road for the connectionist framework, and some arbitration can begin between Gallistel's criticisms of connectionism's computational credentials and the defence to be put forward below. But first, we need to look at the sky.

3.3 Problem Subdomain: Mechanics & Features of the Polarised Sky

Since the pioneering work of Karl von Frisch on the peculiar "dance" of honeybees early in the 20th century, it is well established that many creatures use polarised light in the sky as a navigational cue. Polarisation is caused by scattering of photons from the sun as they collide with particles in Earth's atmosphere. As a result of this scattering, sunlight can become linearly polarised in certain planar orientations. The specific plane in which light is polarised is called the 'e-vector direction'. Polarisation also occurs to various extents, i.e. how much of the light gets linearly polarised in this way. Light is, under one description of course, a transverse wave consisting of electric and magnetic components oscillating orthogonally to one another (see *Fig.* 15). The term 'e-vector' reflects how the *electric* component is the component of light which is being linearly polarised, which means that it is restricted to oscillating within just one plane. Unpolarised light will oscillate in all planes (roughly speaking). It is still a transverse wave, however, and so the oscillating electric component moves up and down in a specific linear plane while also propagating forwards.

As a function of the sun's altitude and azimuthal position (see *Fig.* 16), this scattering of polarised light produces distinctive spatiotemporal patterns of 'e-vectors' arranged in concentric circles across the sky (see *Fig.* 17). This pattern will naturally rotate and change its spatial pattern and distribution according to the celestial position of the sun, but at any one time a given patch of sky will have one direction of e-vector predominating. The rotation and resultant change in the idiosyncratic distribution of any e-vector pattern occurs as a function of time of day, time of year (season), and an observer's position on the Earth. That is, it changes with the local solar ephemeris function. For example, the arc the sun will trace across the sky at summer solstice is very different from the arc it will trace at winter solstice. Similarly, the arcs observed at locations near the equator will be different to arcs observed near the poles. The e-vector distribution is therefore highly dynamic and locally idiosyncratic.
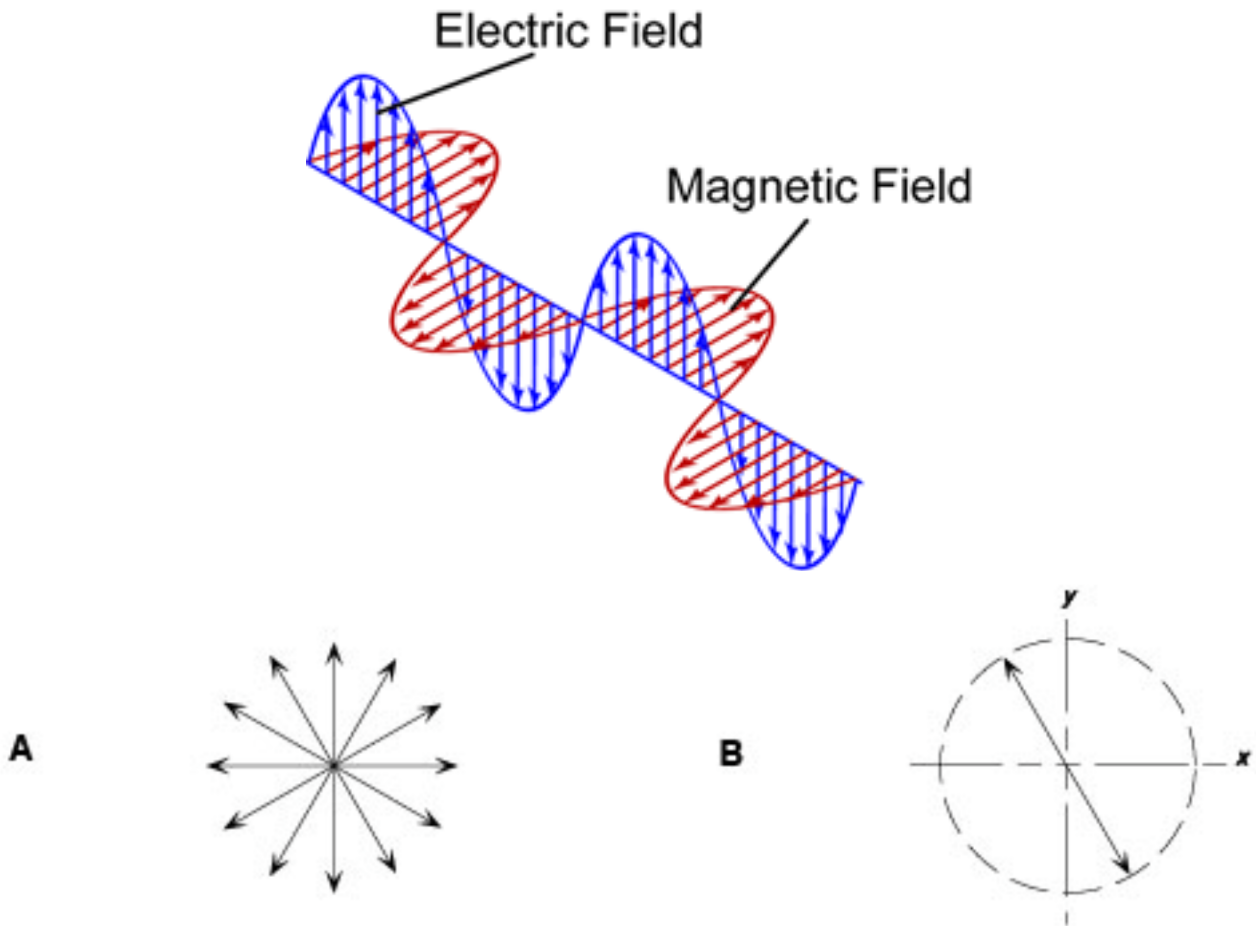
Fig. **15:** Top: graphical representation of light as an oscillating transverse wave. The blue line indicates the electric field oscillating in one plane, the red line indicates the magnetic field oscillating perpendicular plane. Bottom: representations of unpolarised light, **A**, and linearly polarised light, **B**, from front on.
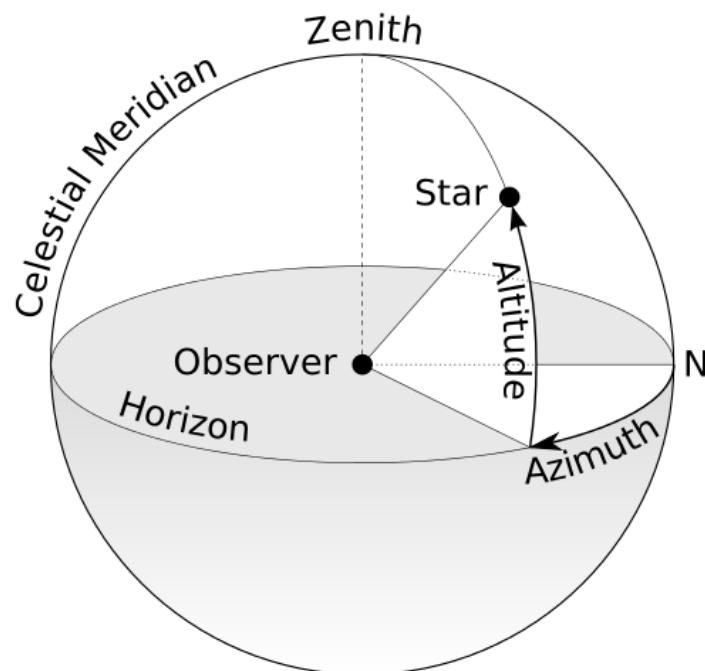

Fig. **16:** Schematic representing notions salient to coordinate systems and navigation. The altitude, or elevation, is the vertical angle away from the horizon of the celestial body in question. Azimuth is the horizontal angle away from a given reference point, in the figure this is shown as north. The zenith defines the point directly above the observer.
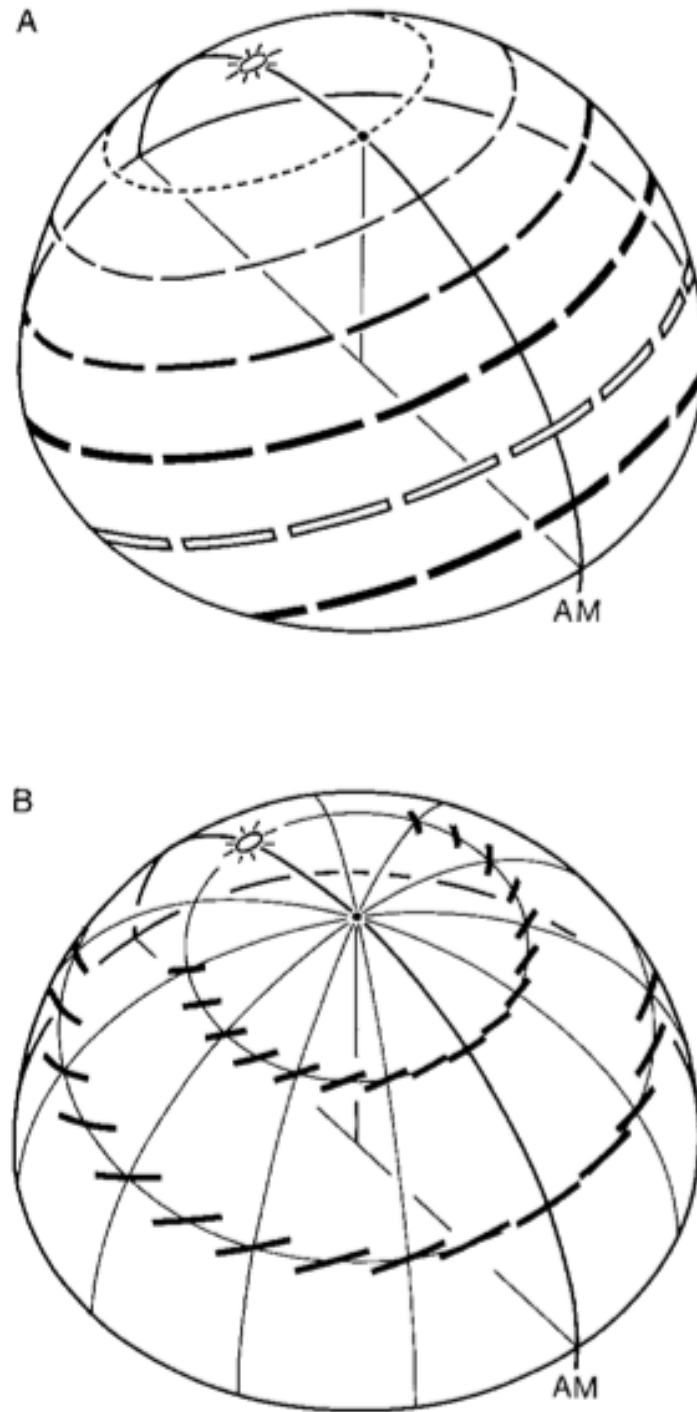
Fig. **17: A** The distribution of polarised e-vectors in the sky, with the sun (open white circle) represented at a solar elevation of 60°. The directions that predominate in any patch of sky are arranged in concentric circles around the sun (or anti-sun). The plane of linear polarisation of a given e-vector is represented by the rotation of the black bars. The degree of polarisation is represented by the thickness of the bars, with the band of maximal polarisation occurring 90° away from the sun (open white bars), and the band of minimal polarisation occurring within ~30° of the sun (dotted line). *AM,* anti-solar meridian. **B** The polarisation patterns for solar elevations of 30° (upper band) and 60° (lower band) placed in a horizontal coordinate system. The sun remains depicted at an elevation of 60°. The <10% polarised e-vectors are omitted, reflecting their functional absence to the creatures. *AM*, anti-solar meridian. (see text for discussion on meridians) (Figure from Rossel & Wehner, 1984: p. 608)

However, there are two spatial properties of the polarisation pattern that are importantly stable. Any idiosyncratic polarisation pattern always maintains two prominent planes of hemispheric symmetry: 1) a 'vertical' plane formed by the great circle of the solar and anti-solar meridians[12], and 2) a 'horizontal' plane formed by the great circle of maximal polarisation that occurs 90° away from the sun (see ***Fig. 17a***). E-vectors are always horizontally polarised along the solar/anti-solar meridians, marking a division between the left/right hemispheres of the sky. The distribution patterns in both 'vertical hemispheres' then mirror each other. Similarly, the hemispheres carved out by the band of maximal polarisation that cuts across the earth, orthogonally to the solar/anti-solar meridians, mirror each other too. However, they do so to a lesser extent for one important reason.

Within a ~30° band around the sun, light is less than 10% polarised, which is below the sensory threshold for biological polarisation detectors. This creates a navigationally salient asymmetry, as e-vectors are effectively absent in the patch of sky that contains and surrounds the sun. The left/right hemispheres still mirror each other in that case, but the solar/anti-solar hemispheres do not. Maximally polarised horizontal e-vectors, therefore, always "point" towards the anti-solar meridian; the azimuthal position that is 180° away from the sun. Performing experiments on the ants either before sunrise or after sunset, or with the sun occluded, demonstrates that they can indeed navigate using polarised light in the absence of the sun (Duelli & Wehner, 1973). However, it is the asymmetry in the distribution of e-vectors due to the sun's position, and the relevant mechanics of polarisation, that is important for navigation, not the sun's position (or absence) itself.

3.4 The Neurobiology of Polarisation Detection & Analysis

Determining just how the creatures capture and use some relevant aspects of polarised skylight was done by experimentally withholding particular spatial or temporal properties of the distribution of e-vectors across the celestial dome. Observing the systematic navigational errors that accrue when animals are artificially exposed to the redacted stimuli then give clues to the computational tasks they perform (Wehner & Rossel, 1985; Rossel & Wehner, 1982; Rossel & Wehner, 1984; Wehner & Müller, 1993). The first kind of behavioural experiment removes spatial components of the pattern of polarised skylight (Wehner & Rossel, 1985; Rossel & Wehner, 1982; Rossel & Wehner, 1984). Ants[13] were first trained to forage at a set feeder location under optimal conditions, i.e. a full blue sky with a complete polarisation pattern available. Once trained to forage at this location, they were then made to recall and internally reproduce the direction of the feeder location. This was operationalised by forcing the

---

[12] There are various ways to define celestial meridians, and much of the literature cited throughout this discussion fails to delineate which is being used, and can be ambiguous across research groups or papers. Given that this literature uses terminology such as, "pointing towards", or "indicating" the solar or anti-solar meridians, it may be helpful to just think of the solar meridian as the half-circle that begins with and includes the sun, and ends 90° away on both sides at the line of maximal polarisation. The anti-solar meridian, which can then be thought of as including the anti-solar section or hemisphere of the sky, is now something that can be properly "pointed" towards, as its azimuthal position will effectively always be asymmetrical. If the anti-solar meridian merely began at horizontal north and south, with the solar meridian being the observable hemisphere of the sky, its azimuthal position would always be ambiguous, and "pointing" towards it would not make much sense at all.

[13] Many of the experiments cited are done on bees, ants, or other insects. Sufficiently similar experiments have also been repeated across many of these species. The navigational behaviour and the underlying neurobiology is sufficiently similar that no particular care will be taken to highlight which experiments correspond to which species, unless directly relevant. Many of the implications are transferable given the conserved morphology, especially between genera within the family of *Hymenoptera*. Locating sufficient biological precedents, in the more general discussion of the theoretical tools available to connectionism, is where the burden of this discussion lies.

trained ants to navigate to that location with only one patch of light containing one e-vector direction presented, with the rest of the distribution filtered out.

The ants make regular and systematic mistakes under these conditions, missing the feeder location by a specific angular amount. Interestingly, the specific amount of angular deviation depended on which specific e-vector direction was presented to them. Experimenters systematically varied the particular e-vector direction presented to the ant to navigate with and noted the error angle they subsequently make for that e-vector. It became clear that the error angle is the same as the angular difference between the azimuthal position the ant must expect that particular e-vector direction to occur in (relative to an internally produced, fictive line of symmetry), and that e-vector's real azimuthal position in the sky (relative to the actual solar meridian) (see *Fig.* **18**).

The natural inference is that the ants must produce some internal "map" of vectors, distributed in a generalised and symmetrical pattern (see *Fig.* **19**). While any actual distribution of e-vectors changes with time of day and year, this static map of vectors demonstrates that the ants assume particular e-vector directions to occur at fixed azimuthal positions, relative to some discernible symmetry plane (e.g., their longitudinal bodily axis). This seems disadvantageous at first, but such a fixed pattern of symmetrically distributed vectors can actually capture some important properties of polarised light.

The neurobiological details underlying this celestial map are also well established, providing a confluence between mechanistic and behavioural data (Wehner & Rossel, 1985; Wehner, 1987; Labhart, 1988; Sakura, Lambrinos, & Labhart, 2008; Labhart & Meyer, 1999). Beginning at the retina, *Hymenoptera* possess what has been referred to as a "matched filter" in their compound eyes (Wehner, 1987). A compound eye is a honeycombed structure, made up of many individual hexagonal lenses, called ommatidia, functioning in concert. Contrast this with the human eye, for example, which is a single lens. In the uppermost dorsal rim area (DRA) of the Hymenopteran compound eye is a functionally specialised group of ommatidia, used solely for the detection of polarised light. This functionally narrow DRA has been shown to be both necessary and sufficient for polarisation navigation (Wehner, 1989b). Within each individual ommatidium of the DRA, two hair-like 'microvillar' UV-photoreceptors are arranged in a cross, perpendicular to each other. UV-photoreceptors are used because they are highly sensitive to light that is linearly polarised parallel to their microvillar direction, and so they activate maximally when exposed to such a stimulus. As the angle between the microvilli directions and the orientation of the polarised light stimulus increases, their firing rate correspondingly decreases, up to angular differences of 90°. After this point, their firing rate begins rising again, as their angle of difference begins to decrease once more.

These first-order detectors thereby exhibit "polarisation-opponent" properties, in that their respective response profiles are sinusoidal functions of the e-vector orientation they are exposed to (Wehner, 2003). Given that at any moment in time each pair will be exposed to the same stimulus, if the 'type-I' microvillar receptor is activating maximally, the opposing 'type-II' receptor will necessarily be activating minimally. The stimulus is necessarily orthogonal to the type-II detector direction if it is parallel with the type-I detector. They are, therefore, polarisation *opponents*.
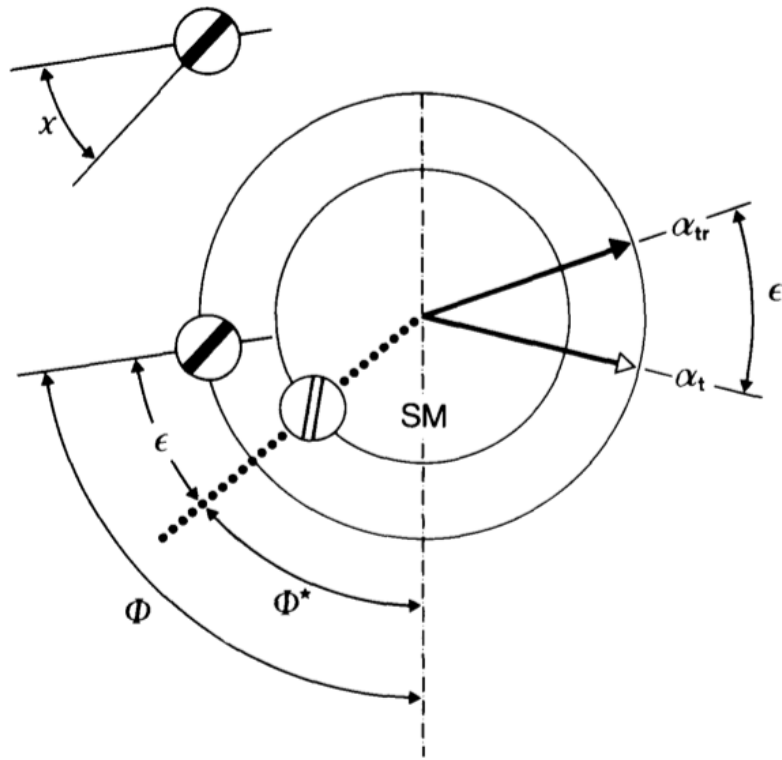
Fig. 18: The experimental design from which the *Hymenopteran* internal e-vector representations were derived. Bees and ants were trained to forage in direction α_{tr} under normal conditions with a full e-vector pattern available. During testing, they were presented with only an individual e-vector, χ (solid black bar), to navigate back to α_{tr} with. In the real sky, χ occurred at azimuthal distance, φ, from the solar meridian, *SM* (outer circle). However, under these experimental conditions the insects navigated towards α_t instead, deviating by error angle ε, demonstrating that the insects incorrectly assume χ to occur at azimuthal distance φ^* (inner circle). φ^* is the same as φ minus ε (open white bar). (Figure from Wehner, 1989a: p. 355)
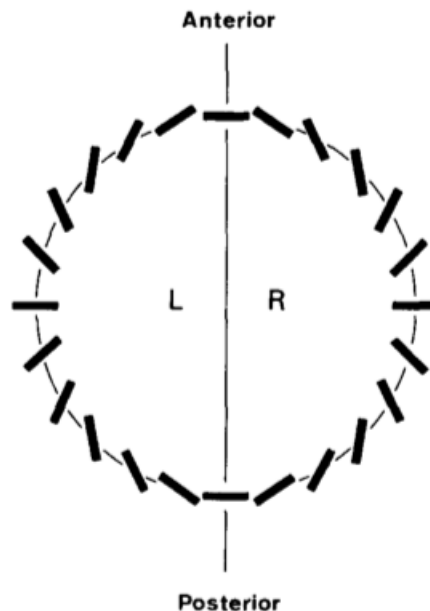
Fig. 19: A geometric representation of the generalised *Hymenopteran* e-vector map, as inferred from behavioural experiment. It shows the fixed azimuthal position they assume a given direction of e-vector will occupy relative to a fixed plane of symmetry. The azimuthal position hypothesised for each e-vector was inferred by the way the creatures regularly and reliably misrepresent their target location when exposed to only that e-vector. *L* and *R* indicate the left and right halves of the visual field, respectively. (Figure from Wehner, 1989a: p. 355)

Having polarisation-opponent receptors creates a greater level of contrast, and thus a greater level of polarisation sensitivity one layer back at a second-order interneuron, or "crossed-analyser" (Labhart, 1988; Wehner, 1989a; 1989b). The second-order interneuron receives antagonistic inputs from the perpendicularly opposed first-order detectors. It receives excitatory input from type-I and inhibitory input from type-II. The integration of those competing orthogonal signals (hence "crossed-analyser") ultimately shapes the second-order interneuron's response profile to be highly sensitive to changes in e-vector direction. It also becomes more finely tuned to the specific angle of orientation that corresponds to the microvillar orientation of the type-I receptors it is connected with (see *Fig.* **20**). The enhanced and refined sensitivity at the second-order neuron also aids in filtering out navigationally irrelevant noise in the signal, e.g. light intensity, that could otherwise disrupt subsequent processing (Wehner, 1989b).

The DRA of the *Cataglyphis* has around 70 of these UV-detecting ommatidia, with their respective type-I receptors oriented and distributed in a fan-like manner (see Wehner, 1987: p. 523). This begins to show why Wehner's description of this sensory system as a "matched filter" is apt: the retinal
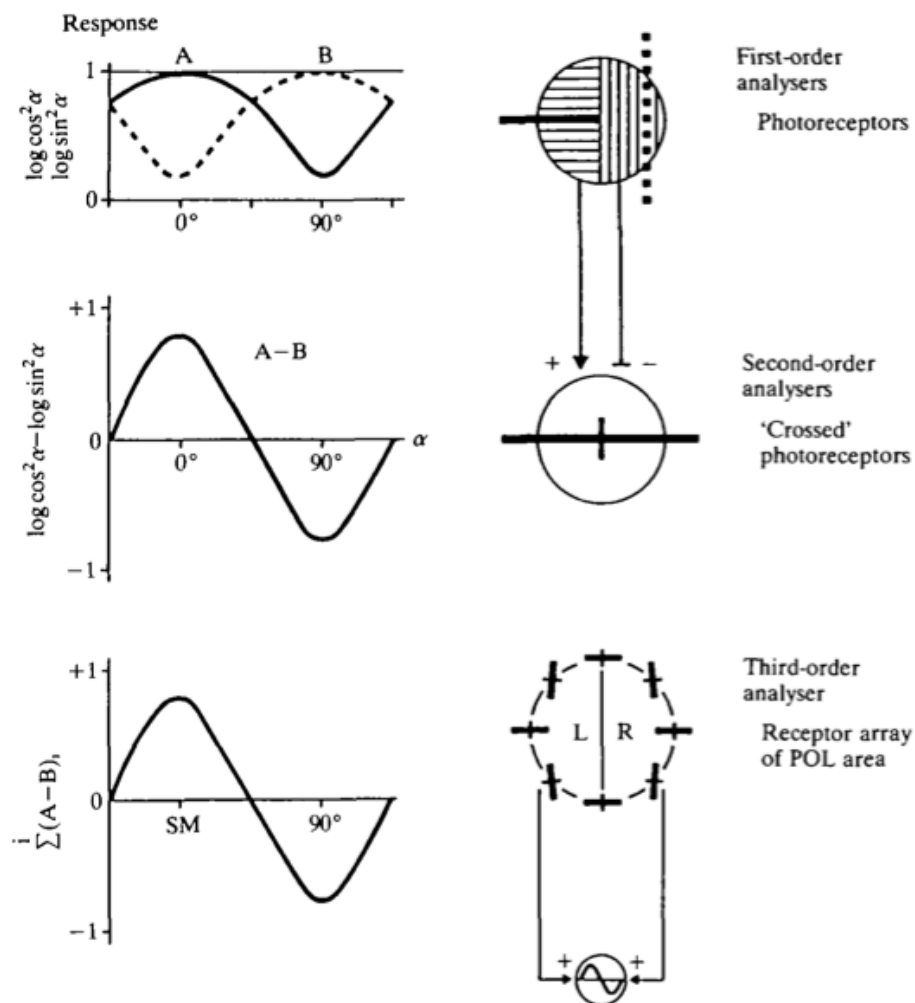

Fig. **20:** Wehner's hypothetical schematic of the successive layers of polarisation detection and analysis in the DRA of the *Hymenopteran*. Upper most figures show the respective microvillar directions of perpendicularly opposed photoreceptors, *A* and *B,* and their response profiles plotted against e-vector direction. Middle figures depict the response shaping effects the antagonistic interaction has one layer back at the 'crossed-analyser' integrating neuron. Bottom figures depict the totality of these neurons functioning as one unified *third-order* compound filter. *SM,* solar meridian. (Figure from Wehner, 1989b: p. 71)

distribution, and overall rate of rotation between these detectors, constitutes what can be thought of as a vector field, one that partially matches the vector field that is the celestial distribution of polarised e-vectors (Wehner, 1987: p. 522). The preferred directions of the retinal detectors in the DRA generally change with the same nonlinear rate of rotation as the static e-vector map that was hypothesised, lending further weight to the claim that the physical geometry of the system is important to the internal vector map.

The matching is a non-arbitrary result of having second-order 'crossed-analysers' each receive inputs from separate populations of the first-order photoreceptors. Pooling information from different groups of these detectors increases their differential sensitivity to particular orientations of e-vector stimuli. As a result of selective pooling, they become "tuned" to more generalised e-vector directions, as in *Fig.* *21* (Wehner, 1987: p. 523; Wehner, 1989a: p. 356; Labhart, 1988). Moreover, a bilateral antagonism between posterior and anterior sections of the DRA's supports the proper differentiation and projection of e-vector directions. If imagining the functional properties of the creatures' second-order integrators as a "circular" and symmetrical map of e-vectors (as in *Fig.* *19*), identical e-vector orientations occur antipodal to each other. By having anterior sections of the DRA send excitatory signals and posterior sections sending inhibitory signals, as the creature sweeps around clockwise it can differentiate
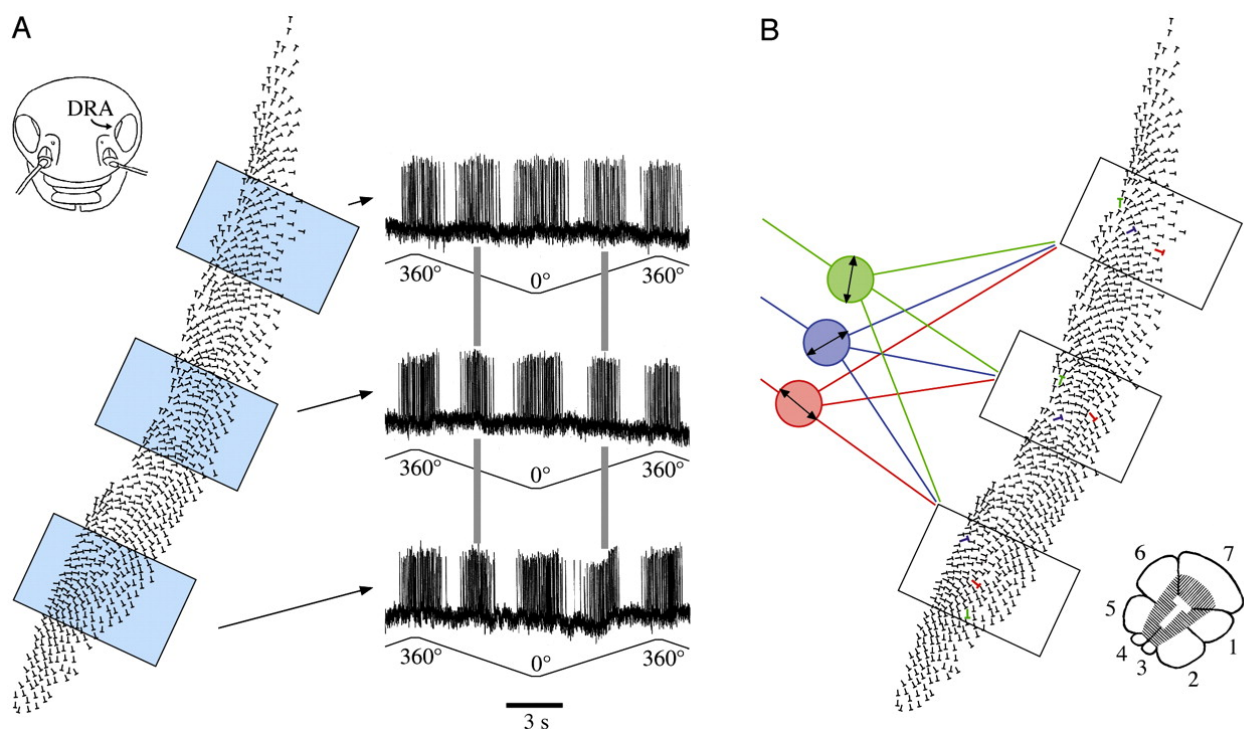

Fig. **21: A** Representation of the response profile of a POL-1 large-field integrative neuron when separate areas of the DRA are stimulated by a blue-light stimulus. Stimulated areas of the DRA shown on the left, their respective spike trains on the right, and the rotation of the stimulus is represented by ascending and descending lines beneath. The neuron responds to all areas of stimulation, maximally firing at the same angle of rotation in all cases, marked by the vertical grey lines. **B** Representation of the interpretative logic of integration in POL-neurons. Three types of large-field integrators and their preferred e-vector orientations are shown by the circles with corresponding double-headed arrows. Each one receives input from many ommatidia across the whole DRA area, but from different kinds depending on their rotation or location (see coloured T symbols). T symbols represent the spatial distribution and orientation of excitatory receptors within the ommatidia, an enlarged cross-section of which is shown inset. (Figure from Labhart, Petzold & Helbling, 2001: p. 2428)

between "front" or "back" projections of identically aligned e-vectors. The same e-vector will necessarily activate each eye in opposite sections (see Wehner, 2003).

The crossed-analysers may also have a complex and pyramidal hierarchy of their own, where many narrow-field integrators pass information to a more limited subset of large-field integrators, and so on. Nonetheless, the spatial distribution and connection profiles of these second-order 'crossed-analysers' forms a neurobiological basis for the symmetrical e-vector map hypothesised by the initial behavioural experiments, as set out in *Fig.* **19** (Wehner, 1989a; Fent, 1986; Rossel & Wehner, 1984). Indeed, Wehner goes as far as saying that, "…the array of second-order analysers *is* the insect's e-vector map" (1989b: p. 73, emphasis in original). Whatever hierarchical idiosyncrasies may exist, e.g. across species and/or genera, a small number of large field integrators, with axes-of-tuning adequately spaced, appears mathematically sufficient to perform satisfactory analyses of celestial polarisation patterns[14]. These large-field integrators are called POL-neurons due to their more specific and refined polarisation sensitivity.

To sum all this up, the ant has built in to its sensory apparatus a 'matched-filter'. A matched-filter is a sensory apparatus, "…in which the spatial layout of a population of receptors is matched to a certain spatial aspect of the navigational problem to be solved" (Wehner, 1987: p. 512). However, the matched-filter must be exploited in some manner in order to determine compass directions for orientation. To do this work, Wehner (1989a) postulates the 'scanning hypothesis'[15].

While navigating, the Desert Ant, "…stops every now and then and performs what looks like a graceful little minuet, a partial or full turn about its vertical body axis" (Wehner & Wehner, 1990: p. 41). The function that this rotatory movement fulfils is locating the symmetry plane carved out by the solar/anti-solar meridians[16]. It achieves this by "sweeping" the matched filter of the DRA across the polarisation pattern in the sky (Dacke, Nordström & Scholtz, 2003). Given the geometric and neurobiological details of the retinal filter discussed above, as the ant completes a rotation, it will produce a wave of summed receptor responses across the second-order interneurons. The sum of the responses from the retinal receptors will reach a maximum value whenever the ant's retina is optimally matched with the current celestial e-vector pattern. That is, best fit is achieved whenever the ant's longitudinal body axis is aligned with the solar/anti-solar meridians. The ant can then use this information as a natural cue: the symmetry plane is a reference angle or calibration point that can be found in any pattern of polarised skylight, at any time. Mechanistically, it will do this by registering the neuron(s) associated with the peak of the time-modulated responses across the second-order analysers. Under normal conditions, this will allow it to align itself with the solar/anti-solar meridian (see *Fig.* **22**).

However, given that the more polarised part of the sky is farthest away from the sun, any horizontally oriented e-vector will necessarily be indicative of the anti-solar meridian (or anti-solar part of the sky), not the entire meridian line. This ensures that even in the absence of the sun, because of cloud-cover or

---

[14] More explanation on why this is the case in Chapter 5.

[15] The scanning hypothesis has been supported by other experimental results, the details of which are beyond the scope of discussion here (see Rossel & Wehner, 1986).

[16] See **3.3** & *Fig.* **17** for a recap on meridians.
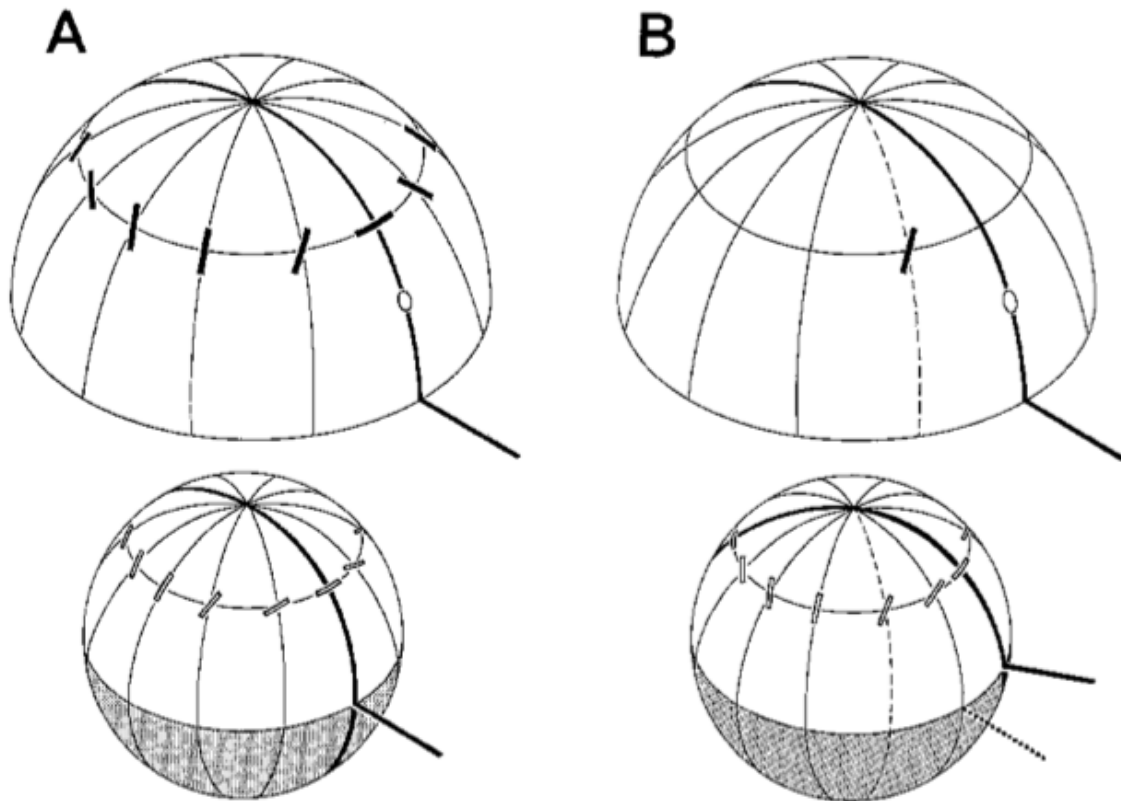
Fig.* 22:** Idealised representation of the mechanics involved in scanning the polarisation filter across the sky. Upper figures show the celestial dome, with full polarisation pattern available (A***) or a single e-vector in one patch of sky (***B***), for one given latitude. The black bars represent the direction of e-vector. Lower figures show the *Hymenopteran* vector map/ polarisation filter in the DRA, with microvillar/tuning directions shown by the open white bars. Solid black lines represent the symmetry plane of the skylight (upper) and longitudinal body axis of the creature (lower). In ***A*** best fit is achieved when both planes of symmetry coincide. In ***B***, best fit is only achieved after the animals deviate from the actual symmetry plane of the sky. In this case, deviating by 30° to the left is required, demonstrating how the subsequent navigational errors accrue from initial inaccuracies in calibration. (Figure from Wehner, 1987: p. 522)

experimental intervention, maximal activation will occur when the ant is aligned towards the anti-solar section of the sky specifically, and can use that as its calibration point. With a necessary bias towards seeking out the anti-solar meridian, and setting that as the reference angle, it can be used to define the azimuthal positions of all other directions. This exploits the unique spatial asymmetry of polarisation scattering. Namely, it exploits the effective lack of any such scattering within a ~30° band around the sun[17] (Rossel & Wehner, 1982).

This completes the story as to why we observe the navigational errors that initially led to the internal "vector map" hypothesis, and the subsequent postulation of the scanning hypothesis: when only one e-vector direction is available for the ant to process, the relevant neural dynamics during rotatory

---

[17] How the creatures discriminate between the solar and anti-solar meridians is thought to also involve parallel processes that analyse differences in spectral information as well as degrees of polarisation that arise at those points (Wehner, 1989a: p. 356; Wehner, 1989b: p. 77; Fent, 1986; Rossel & Wehner, 1984). For example, the ratio between light of long wavelengths (460–700 nm) and light of short wavelengths (300–460 nm) is higher in the solar hemisphere than in the anti-solar hemisphere of the daytime sky (el Jundi, Pfeiffer & Hoffberg, 2011: p. 9). *Cataglyphis* ants have been demonstrated to univocally orient towards a 90° (horizontal) vector as if it were indicative of the anti-solar meridian, and successfully navigate even at times near sunset or sunrise, when the polarisation pattern is ambiguous, but the spectral and intensity gradients are not. Spectral gradients are thus operating as an effective redundancy when the preferred source of navigational information is unsatisfactory or unavailable (Fent, 1986).

scanning movements are interrupted, such that the maximal value of summed receptor responses is shifted towards where that e-vector direction is located within the spatial schema of the internal vector map. The ant ends up aligning itself incorrectly in the world, towards where that single e-vector direction is in relation to its own internal, 'fictive' anti-solar meridian, rather than where that e-vector is positioned relative to the veridical anti-solar meridian. The error angle displayed is commensurate with the neurobiological details of the celestial map, and the behavioural evidence for the scanning hypothesis (*Fig.* **18** & **22**). Note that it is not the case that the presentation of single e-vectors only activate those detectors which match its orientation perfectly. All the detectors get activated to some extent during scanning movements (unless and until they are orthogonal to the e-vector stimulus). It is their variable activation rates taken *together* — caused by their greater or lesser directional match with the e-vectors in the sky, in concert with clever wiring and distribution — that determines how the creature eventually aligns itself in the world.

Other insects similarly perform spinning movements, and are hypothesised to do so for the same calibration reasons, e.g. dung beetles (Dacke, Nordström & Scholtz, 2003: p. 1540). The scanning hypothesis is thus well supported empirically, with verification coming from various related species.

3.5 Tracking the Moving Sun

While a confluence of neurobiological and behavioural evidence for both the scanning hypothesis and the matched filter is very compelling, such an elegant but economical mechanism naturally has its disadvantages. It inevitably leads to some error rate, even under normal conditions. An important gap in the discussion so far has been a focus on deriving compass information *synchronically* from a static pattern of polarised skylight. However, the azimuthal position of the sun obviously moves as time passes. By implication, the calibration point the ant uses, the anti-solar meridian, changes with it. At first glance, this may not be so disastrous, as not much time actually passes between the foraging and return legs of the ant's navigational task. This is particularly true in the case of *Cataglyphis bombycina*, for example. As discussed earlier, the heat based time constraints limit foraging to around ten minutes, and even at local noon when the sun's azimuthal rate of change is highest, it doesn't move all that far. This minimises navigational error, making the matched filter sufficiently adapted to the creatures' environment.

However, ants do seem to possess some "knowledge" of the long term movements, and even more interestingly, that *changing rates* of movement, of the sun across the sky. That is, not only do they display some knowledge of how the morning and evening polarisation patterns are 180° mirror images of each other, but also that the sun's azimuthal rate of change is nonlinear: the higher the elevation of the sun, the smaller the circle its trajectory traces around the Earth. The sun thus completes more rotational change in the same increments of time when at higher angles of elevation.

This leads to the second type of experimental setup, which omits temporal aspects of the polarisation pattern in the sky. Several experiments took *Cataglyphis* ants and restricted them to only seeing the sun and sky for a set period of time. This was done just as they were beginning their foraging lives[18]

---

[18] Forager ants come in identifiable cohorts, with rapid replacement and turnover rates, and a foraging lifespan of around 5 days (Wehner,1993).

(Wehner & Lanfranconi, 1981; Wehner & Müller, 1993). Some were restricted to seeing the sky only for a few hours in the early morning, around dawn to 9:30am solar time (Wehner & Müller, 1993). Other groups were restricted to only viewing the sky around local noon[19] (Wehner & Lanfranconi. 1981). Both experimental groups were allowed to forage under these conditions normally, but then caught and kept in the dark at the close of the experimental window. They were then reintroduced to their colony after sunset by the experimenters. Thus, any knowledge the ants could derive directly from experience about the sun's azimuthal positions and rates of movement across the sky was severely limited. This procedure was repeated for several days, until on the final day they were again caught and kept in the dark, but released into a testing area at various later times to navigate back to the colony by themselves. Having had no prior experience navigating at these times, this procedure allowed for clear demonstrations of what the ants knew about the sun's movement, and how accurately they inferred such movement from very limited experience, if at all.

The results are staggering given the relative paucity of information available during the training phase. Ants that were allowed to forage in the morning and then tested in the afternoon clearly expected the sun to have moved ~180° from where it was when they had been exposed to it (***Fig.* 23**). Similarly, ants that were trained and exposed at around noon accounted correctly for the slowing rate of azimuthal change when released and tested in the evening (see Figure 2 in Wehner & Lanfranconi, 1981: p. 732). A similar but slightly adjusted procedure that tested ants only an hour after their training phases produced a more fine-grained result, giving us a combined picture that the ants' internal representation of the azimuthal rate of change of the sun is a loose generalisation of the real sun-azimuth curve (***Fig.* 24**). These effects were cancelled out in the longer procedures where underestimations were roughly counterbalanced by overestimations across trials, and vice versa.

Even though they may overestimate the lowest rates of movement of the sun's azimuthal position and underestimate the highest rates, we can clearly see that their expectations resemble the general relationship between time and azimuthal rate of change. It picks up rapidly and peaks with local noon, then drops off markedly at first, with a slowing rate of decline after that, thus approximating the nonlinear rate of change of the real sun-azimuth curve.

Two hypotheses can be dismissed in light of these data: the 'interpolation' and 'extrapolation' hypotheses (see ***Fig.* 23**) (Wehner & Müller, 1993). The former assumes the ant linearly interpolates between the prior positions of the sun that it was exposed to and then extends that interpolated function into the periods when it was isolated in the dark. The latter assumes a linear extrapolation from the most recently observed rate of movement of the sun to its estimated current position. As in ***Fig.* 23** (lines ***B*** and ***C***), the predictions of these hypotheses are demonstrably inconsistent with the behavioural data. Moreover, they would be poorly adapted functions for approximating the actual local solar ephemeris function, were the ants to use either of them. Wehner and Müller (1993: p. 333), on the basis of these results, conclude that, "…*Cataglyphis* is informed innately about one general spatiotemporal aspect of the sun's 24-h course, namely, that the angular position of the solar meridian at sunrise and sunset lie opposite to each other".

---

[19] Local noon is whenever the sun is at its highest, which may obviously differ from 12:00pm local time.
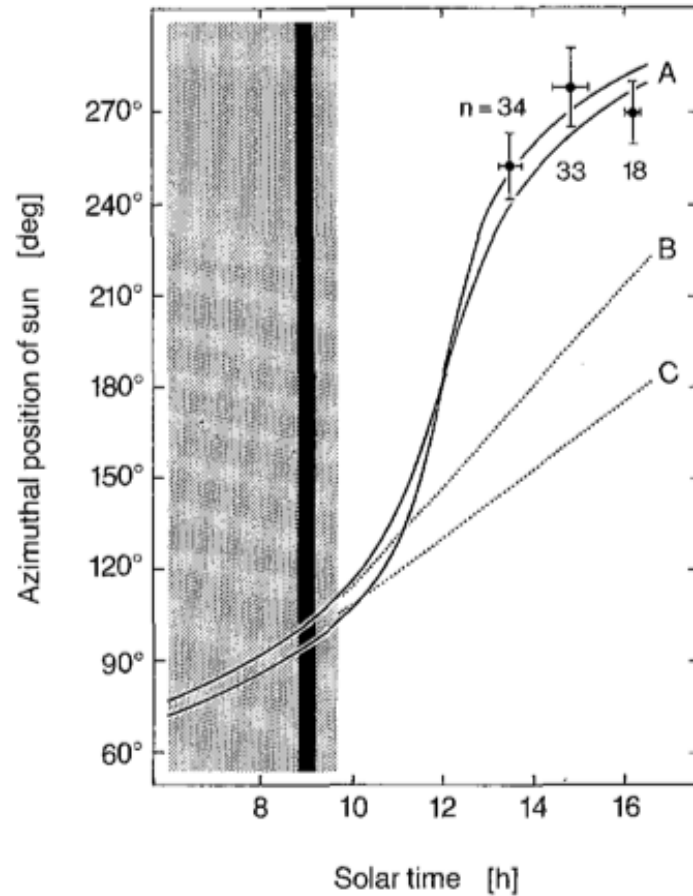
Fig. 23: Azimuthal positions of the sun in the afternoon as assumed by ants, *Cataglyphis fortis*, who only had prior experience of its location and movements in the early morning (lightly shaded area). The black bar indicates the training time, during which they were trained to forage at a particular location (~9am). Data points show the positions separate groups assumed the sun to have moved to, as deduced from their homing behaviour during testing (groups were kept in the dark for ~4:32, ~5:54, and ~7:22 hours, respectively). *A*,
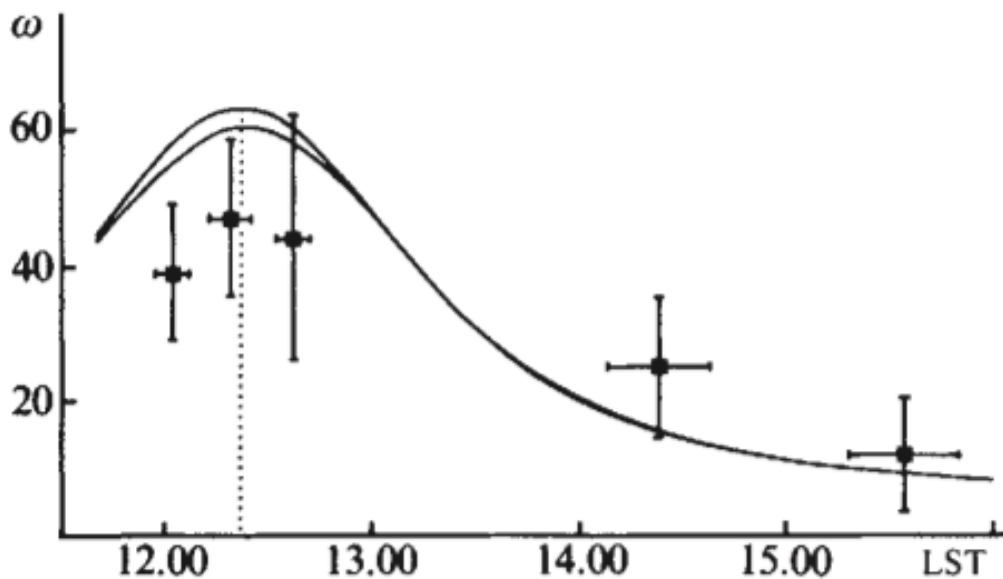
Fig. 24: Errors made by ants, *Cataglyphis bicolor*, while trying to compensate for the daily azimuthal movement of the sun. Solid lines represent the actual sun-azimuth curve for the times and location of experiment. ω, change in azimuth position per hour (deg h^{-1}). *LST*, local solar time. Broken line, local noon. The data points indicate the ants' estimates of the rate of change in the sun's azimuthal position during the time that they were caught and restricted by experimenters, as deduced from their subsequent homeward courses. Details in text. (Figure from Wehner & Lanfranconi, 1981: p. 733)

However, perhaps the ant's representation of the solar ephemeris function is filled in with more detail on the basis of local experience. This is tentatively supported by experimental data where ants that were tested at local noon failed to accurately predict the *sign* of the ephemeris function. That is, they estimated that the sun would move west-to-east, rather than east-to-west, over the course of a day (Müller, 1987: cited in Wehner & Müller, 1993). The ants that guessed wrongly subsequently navigated in the complete opposite direction than required. More recent work has added to this hypothesis that learning through experience can play an integral role in path integrating animals. In particular, an effect has been observed for more junior members of foraging cohorts (Grob, Fleischmann, Grübel, Wehner, & Rössler, 2017).

Central place foragers, like Desert Ants, use other cues like smell, and panoramic view snapshots, to locate their nest when coming into close proximity to it at the end-stages of their return trip. As ants transition from interior colony worker to exterior hunter, they perform practice runs to learn these cues by performing looping walks around their nest over the course of three days (Grob et. al., 2017). During these learning walks, they periodically make sharp turns back in the direction of their nest. This behaviour is presumed to be part of forming snapshot memories to calibrate with their external, celestial reference system. Experimental tests systematically altered or removed elements of celestial information for some groups of ants performing these learning walks, and yet these groups were still able to accurately pirouette to the location of their nest (Grob et. al., 2017). The initial reference system that aligns their backwards gazing to the nest entrance is therefore independent of the polarisation pattern and celestial compass system.

However, neuroanatomical analyses showed that only the learners who were exposed fully to a naturally rotating polarisation pattern had significant volumetric increases to a number of centrally located brain regions that receive projections from the DRA and other visual-sensory arrays. These brain regions, e.g. the central complex and mushroom bodies, are highly correlated with long-term memory formation, learning, and sensory integration. So, while polarised light can't be the reference system used for their learning walks, the *moving* polarisation pattern is necessary to trigger neuroplastic events in regions correlated with integration of the celestial navigation system.

Ultimately, an innate general template that lets the ant estimate how the morning and evening positions of the sun are roughly mirrored says nothing about the important stages in-between. As mentioned earlier, the specific trajectory the sun takes across the sky is locally unique to geographical location as well as time of year. The sun takes a low and short arc at winter solstice, a higher and longer arc at summer solstice, and a roughly identical one all year round if you're located at the equator. Given that *Cataglyphis bombycina* only forages at a time of day when the sun's rate of azimuthal change is highest, and is not located at the equator, it seems reasonable that there would be *some* selection pressure to become informed, or capable of *learning* about these local details. Moreover, as the Grob et. al. study suggests, only the naturally moving polarisation pattern engages a number of neuroplastic changes to brain areas related to learning and memory. There is *something* being learnt about time, or about something that is itself sensitive to time.

The upshot of this final section is that the data at hand suggest three conclusions. First, there is a gap in the built-in knowledge of Desert Ants, where the creatures need to learn some specific details about the

local ephemeris function through experience. Secondly, it suggests that there can, at least some of the time, be a lack of such information that can negatively impact navigational success (e.g. guessing the sun's movement will be west-to-east). Finally, and importantly, any process of refinement of the general template to more closely fit the local ephemeris function must be reconciled with how they are innately informed about those more general properties of solar ephemerides and polarisation patterns.

To boil it down, we must go on to ask: what aspects of the local solar ephemeris function do the ants actually learn, if any, and by what mechanism? However, as forewarned in **2.5**, how we first understand and interpret the manner in which Desert Ants collect and represent their sensory data, the data that can inform a computational mechanism that performs path integration, will significantly affect how *we* theorists go on to frame this problem regarding the Desert Ant navigational system, and its system for *learning* and adapting to time. Before we can tackle the problem of learning, we must fully understand their representational schema, and computational framing, first.

Chapter 4

The Explanatory Demands of Desert Ant Navigation

Conceptual Limits & Practical Challenges

"How could one encode a number using changes in synaptic conductances?"

- *C. Randy Gallistel & Adam King*

4.1 Gallistel on the Constraints of Computation

Learning the local ephemeris function represents a distinct challenge. Even though any natural foraging excursion by a member of the *Cataglyphis* family is short enough that such knowledge of the ephemeris function seems superfluous, they nonetheless demonstrate an amazing ability to account for it when artificially forced to. The challenge is:

> Thus in order for the ant to use polarized skylight as a reliable source of directionality information, it must in some sense 'know' how patterns of skylight polarization correlate with time of day in the local environment; that is, *it must know the local ephemeris function.* […]
>
> To register such correlations, the insect brain clearly must have some way of matching information about skylight polarization patterns with information about time of day. (Morgan, 2014: p. 69, emphasis added)

The idea that computation is the driving force behind successful navigation returns to the fore here, as a conceptually rich explanatory framework for understanding just how the ants could correlate particular polarisation patterns with particular times of day. Put another way, the ants could use bread-and-butter computational processes to encode, link together, and recall values for polarisation patterns and time. In doing so, they would be building a familiar kind of malleable but faithful data structure, as in **2.2**, that informs them about the local ephemeris.

However, this computational approach has a couple of obvious implications. First, organisms must be sensitive to the kinds of information required to perform that computation. Second, such information must be preserved over time with adequate fidelity. If information storage is leaky, noisy, or otherwise "lo-fi", computation will produce misleading results. Only an organism that reliably registers and stores information is capable of performing those tasks that fall under the cognitive umbrella. In other words, cognition requires the ability to learn and to memorise — two distinct but necessary capacities: "learning is the extraction from experience of behaviorally useful information, while memory is the mechanism by which information is carried forward in time in a computationally accessible form" (Gallistel & King, 2010, p.279). For the Desert Ant, it must learn and memorise the local ephemeris function.

Gallistel argues that, "[all] of this presupposes a symbolic memory, a memory capable of encoding information about the world and the animal's relation to it and carrying that information forward in time in a computationally accessible form" (2008: p, 234; cited in Morgan, 2014: p. 71). Moreover, Gallistel believes that having a classical symbolic read/write mechanism is, "…*essential* to computation and representation" (Gallistel & King, 2010: p. 230, emphasis added). Gallistel believes that there can be no real-valued representations of environmental variables without a symbolic read/write head. By implication, only an organism with a symbolic read/write mechanism is capable of completing a task such as path integration and dead reckoning.

So, for Gallistel, successful dead reckoning relies on the ability to store symbolic representations of values for relevant variables, e.g. 'distance' and 'direction' of travel. Subsequently, there must be a

mechanism to retrieve and manipulate those symbols to compute a return path. In short, dead reckoning requires a system capable of calculating with numbers. This is where connectionism runs into trouble, says Gallistel. He claims that connectionist networks have no reliable means of storing numbers (Gallistel & King, 2010: p. 279), and given how computation allegedly works, this implies that connectionist networks are powerless to compute the trigonometric functions required for dead reckoning.

The problem for connectionism is distilled neatly into a single piercing question, "[how] could one encode a *number* using changes in synaptic conductances?" (Gallistel & King, 2010: 279, emphasis added). In this quote, Gallistel is referring to the orthodox Hebbian (Hebb, 1949) framework of learning and memory formation in the cognitive and neurosciences, i.e. the "neurons that fire together, wire together" principle. Gallistel is right in asserting that connectionist learning and representation relies heavily on the Hebbian model of synaptic strength adjustment (see **2.3**).

However, if connectionism's primary conceptual tool cannot sustain symbolic representations with which to learn and store the relevant information, then it can't compute the requisite trigonometric functions to achieve path integration. Gallistel and King (2010) make these arguments concrete by demonstrating how they apply to a neural network model of path integration developed by Samsonovich and McNaughton (SMc) (1997: p. 5902) (*Fig.* 25).

First, the SMc model is computationally expensive, to the point of biological implausibility. It works by using a rectangular array of neurons to constantly integrate various pieces of information, like 'speed' and 'direction'. The integration process creates a localised "bump" of activity within that array,
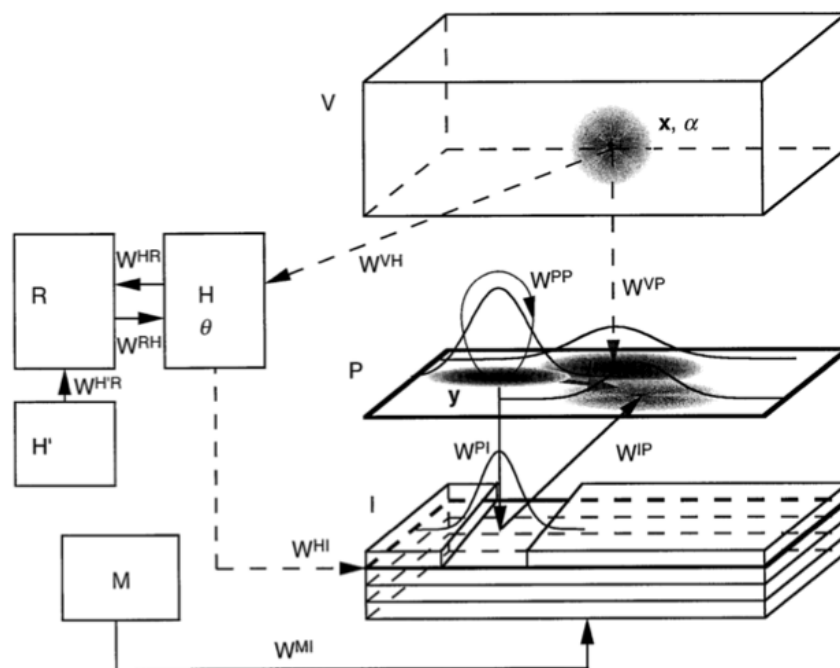

Fig. **25:** Samsonovich and McNaughton (1997: 5902) Attractor Network Model of path integration in the rodent hippocampus. A 'bump' of reverberating neural activity in the *P*-array of neurons is understood as a self-sustaining attractor. The bump and its movements estimate absolute position across the planar environment. The bump is created by integrating information about heading from *H*, with efferent signals from some motor pathways, *M*.

sustained by reverberatory loops. The movements of the bump of activity across the planar array of neurons represents the absolute position of the animal traversing the two-dimensional floor of its environment (Samsonovich & McNaughton, 1997). This central array alone, however, has more than two billion synaptic connections, and the entire SMc model has around one million model-neurons. Unfortunately, the ant brain has somewhere between a quarter of a million and a million neurons. Even with a million neurons, presumably the ant brain does much more than just path integration. The SMc model costs more in computational resources than biology can afford for the Desert Ant.

The second issue is the SMc model's use of reverberatory attractor dynamics to sustain the 'bump' of activity. The problem here is that while Gallistel and King concede that a reverberatory loop is the only mechanism in the SMc model that could plausibly implement a symbolic read/write function, they don't think it makes for a very good one (see Morgan, 2014: p. 195-197 for review).

Information contained in a dynamic, reverberating loop is highly volatile. It takes an extremely specific set of target conditions to get the attractor bump going in the first place, and when it is up and running, the level of noise present in the recurrent connections will affect how long it will remain stable. Each time the loop goes around you need only lose a small amount of fidelity in the signal due to biophysical noise for it to rapidly amplify into massive amounts of information corruption in subsequent passes.

For the representational strategy of the SMc model, this is disastrous. The SMc model uses an entire array of neurons to encode a single variable, i.e. the animal's integrated location. The symbolic value for that encoded variable is given by the position of the reverberatory bump in the central planar array. This is a 'place coding', and therefore any noise induced corruptions in the shape and position of the bump will lead to misrepresentations of location.

As it happens, the SMc model puts in place a number of idealisations, one of which serves to effectively eliminate the presence of noise (Morgan, 2014: p. 187-188). Even with this idealisation in place, the SMc model only reports stable activations over 6 seconds of simulation. Gallistel and King (2010: p. 262) duly point out that Desert Ant expeditions can last into the tens of minutes.

Gallistel and King see this as demonstrating the conceptual problems inherent in connectionism. All connectionist networks are burdened by the lack of a faithful symbolic read/write mechanism, which is necessary to store and compute over the values required for intelligent action. Connectionist networks pay the computational price by requiring implausible amounts of complexity, as demonstrated through the SMc model. Even with exorbitant resources, the stability and resilience to noise of reverberating loops remains unproven over the lengths of time required. Connectionism is therefore doomed as an explanatory programme in cognitive science, as it can't even handle a rudimentary example of intelligent action like that of the Desert Ant. If we call the brain a computational device, and our model of the brain's computational hardware lacks the central and defining feature of computing machines, then that model can be discarded.

4.2 The Straw-Man & Synapse Adjustment Reconsidered

The problem with the above criticism from Gallistel is that it 1) makes no reference to the analog framework for understanding connectionism, and 2) is an incomplete analysis of the task domain. The way the problem is pitched is, naturally, a consequence of Gallistel's theoretical commitments. Gallistel believes only symbols, disciplined by a read/write mechanism, can be made to sustain a functioning homomorphism with some relevant quantities in the ant's environment. And only functional homomorphisms can underwrite knowledge structures rich and flexible enough to inform behaviours such as path integration.

First, this is a straw-man of connectionism. Analog computers are distinct machines with a distinct, non-symbolic representational medium (see **1.3 & 2.3**). They use structural representations, and thus don't create symbolic data structures. So, when Gallistel criticises the SMc model for not instantiating a faithful read/write mechanism with which to create a symbolic data structure that matches times of day with patterns of skylight, this completely misses the point of connectionism. Moreover, the SMc model is decades old. Even if it is "representative of a large class of models" (Gallistel & King, 2010: p. 265), it is not the *only* class of connectionist models that tackle the problem of integrating location.

Gallistel may want to argue that symbolic representations are superior to structural representations. However, that argument is never made. Symbolic representations, functional homomorphisms, and classical computational mechanisms like read/write heads are all taken as the *only* possible options on the menu. Gallistel puts a lot of theoretical work into substantiating the usefulness of these classical concepts, of course, they're not merely asserted without justification. Using these concepts to measure connectionism's theoretical viability is still a straw-man though if connectionism fundamentally eschews these notions for some alternatives. The alternatives must be met on their own merit, as the usefulness of these alternatives is also substantiated in a vast body of literature. Ignorance of this literature and how it places the notions of structural representation, structural homomorphism, and analog computation in an independent framework cannot save Gallistel from being hoist by his own argumentative-petard. The strong and explicit claims that 1) the behavioural data on Desert Ant navigation "presupposes a symbolic memory", and 2) that symbolic read/write mechanisms are "essential" to computation, are both false. Symbolic read/write mechanisms are essential to *digital* computation only. The assertion that the data necessarily presupposes a symbolic representational strategy, and that the SMc model is representative of how *all* connectionist networks could ever tackle problems of location and navigation, leads to the second problem.

The second problem is that Gallistel makes a parochial analysis of the problem domain. Parochial analyses of a problem result in parochial framings of the possible engineering solutions therein. As introduced in **2.5**, different methods are available for representing values for the same kinds of environmental variables. These various representational strategies affect the framing of the problem at higher resolutions as a function of their own idiosyncrasies. You create different systems altogether, like Cartesian vs Polar coordinate systems, depending on the particular representational medium — and they require different methods of manipulation.

Gallistel is wedded to one particular strategy, i.e. that, "…there must be an addressable read/write memory mechanism in brains that encodes information received by the brain into symbols…and transports it to computational machinery that makes productive use of the information" (Gallistel & King, 2010: p. *viii*). As a result, he is actively *looking* to frame the problem of Desert Ant navigation in a symbolic fashion. That is, he is working on the assumption that the *only sensible way* to computationally frame the problem is that the Desert Ant *must be* explicitly representing 1) idiosyncratic patterns of polarised skylight, and 2) times of day. This is because the skylight patterns change with time. If the ant is using the skylight pattern for orientation, the ant must update and take into account its dynamic nature. The ant therefore must cross-check a stored almanac that matches values from 1) with values from 2) when faced with some real pattern of skylight during navigation. Thus, the only way to make values of a polarisation pattern reliable for navigation is to have them time-matched:

> …the ant's brain must record specific registrations between polarization patterns and times, and carry these forward in time, so that they may be integrated into a running estimate of the ephemeris. (Morgan, 2014: p. 71)

The only sensible way to represent and match variables like 'times of day' and 'patterns of skylight' in an addressable almanac is to have them represented symbolically. These time-matched patterns, tabulated symbolically, thus explicitly represent the ant's learned estimate of the local solar ephemeris function. Such a symbolic data-structure can be stored and updated, then withdrawn and operated over when necessary in order to successfully dead reckon.

However, section **2.3** showed that there is an independent representational medium available for cognitive scientists to appeal to, which can in turn suggest a different computational logic for solving the problem. First, what if the ant is *not* using polarisation patterns for orientation in the explicit manner that Gallistel suggests? Second, if it isn't, we must reconsider the theoretical significance of synapses and their malleable weights as the method of updating any knowledge-structure the ant in fact possesses. This is because representation is, in an important respect, prior to learning. One's theory of representation will inform one's notions of learning, because different representations require different methods of changing or updating their contents. And what is learning, if not changing or updating the contents of one's representations of the outside world?

There is universal acceptance that information extraction from experience is mediated by synapse adjustment. But what is the theoretical significance of this biological process? Gallistel says changes in synaptic connectivity cannot be interpreted as the store of information extracted from experience which informs later behaviour. This is because synapse adjustment cannot instantiate and store symbols. He writes, "[Synapse adjustment] may explain reflex modification phenomena, but it does not explain the learning of behaviorally important facts and the formation of data structures" (Gallistel & Matzel, 2013: p.172). Using a structural representational interpretation, this can be rehashed.

'Connection weight representation' has been postulated as the second and complementary form of structural representation in connectionist models, next to activation patterns and their abstract geometry (O'Brien & Opie, 2005). The overall pattern of weights modulating input between layers of

nodes can themselves be interpreted as sustaining a resemblance relation with some aspect of their task domains. Think of the orrery from Chapter 1 once more. The internal geometry created by the spheres' positions, at any snapshot in time, sustains a structural resemblance relation with the real planets, which makes the orrery useful as a model of their relative orbits. What sustains that geometry though is the set of variously sized *gears*, i.e. the "connection weights" between the representing spheres.

In virtue of the gears having their own appropriate geometrical structure, their causal interactions which govern the rotations of the spheres *resemble* the gravitational interactions between the celestial bodies and how they create variously sized elliptical orbits. Analogicity is at work then in the connective tissue of the orrery. We are licensed in interpreting the gears as representational vehicles, and interpreting them as structural representations specifically, because it is their relational geometry that sustains a physical analogy via structural homomorphism with an aspect of their task domain.

So, when Gallistel asks *how* a "number" could be encoded in a synapse, the connectionist can reply, "*why* would I need to?". Connection weights, under a proper connectionist interpretation, should be considered as elements in a represent*ing* set of physical magnitudes, whereby some physical relations between these elements track, and can be mapped onto, some physical relations between elements in the environment, the represent*ed* set. Moreover, this mapping of relations between sets of elements is useful and productive to some user of the represent*ing* set of elements, insofar as the user is brought into some contact with the represent*ed* set of elements in the world as a result. They are structurally homomorphic, not functionally homomorphic, to their represented domain.

Reinterpreting neural firings and their connection weights as non-symbolic thus refreshes the way we can interpret how the Desert Ant is representing and making use of the polarised skylight patterns. Chapter 3 demonstrated that they must know something about time and its relation to the sun's variable arc across the sky. However, if they're not creating a symbolic almanac, exactly what kind of knowledge structure are they building and then leveraging towards the task of path integration and dead reckoning?

The next chapter will look in detail at how a connectionist can invoke structural representations towards an analog computational model of Desert Ant navigation and path integration. It will do so in a way that respects and draws on the neurobiological and behavioural data analysed in Chapter 3. It will also involve reframing the problem away from the "obvious" (and perhaps anthropocentric) representational strategy of explicitly representing real idiosyncratic patterns of skylight and time-matching them in an almanac. Instead, it will look at how assemblies of neurons, and the connective structures that modulate their interactions, can sustain physical analogies with certain geometrical aspects of patterns of skylight and traversed paths.

Chapter 5

Defending the Representational & Computational Credentials of Connectionism

Seeing & Solving: Through an Analog Lens

"[The ant's] trick is to incorporate the fundamental spatial properties of the navigational problem into the very periphery of its nervous system, into the spatial design of its sensory surface… and there, algebra gives way to geometry…"

- Rüdiger Wehner

5.1 Preliminaries: Vectors & Trigonometry in Path Integration

The relevance of vectors and their relation to various trigonometric functions, and of trigonometry in general to the problem of path integration, must be understood properly before we dive headlong into a discussion of neural analogs of these concepts in the connectionist tradition. I will demonstrate this using the simplified example from *Fig.* **14**.

We're given two vectors, representing two outbound paths taken: V_1 at 150° with a length of 1m, and V_2 at 50° with length 2m. The dotted line in *Fig.* **14** is the unknown global vector we need to solve for. The digital, algebraic method through which we humans attack this is "component analysis" in a Cartesian plane. This method exploits the fact that any one vector is describable by two other vectors, i.e. its *components*. By breaking each individual local vector down into its X and Y components, depicted in *Fig.* **26a**, we can perform some simple arithmetic (*Eq.* **1**) to get a new set of summed X and Y components that relate to the global vector. These components complete two-thirds of a right triangle, depicted in *Fig.* **26b**.

By following the formulae depicted in *Eq.***1**, the total displacement across the two axes is summed. Here we begin to see the relevance of trigonometric operations to the target behaviour. The *sine* function gives the vertical Y value, with positive values in the upper quadrants, negative values below. The *cosine* gives the horizontal X value, positive values being in the right quadrants, negative in the left.
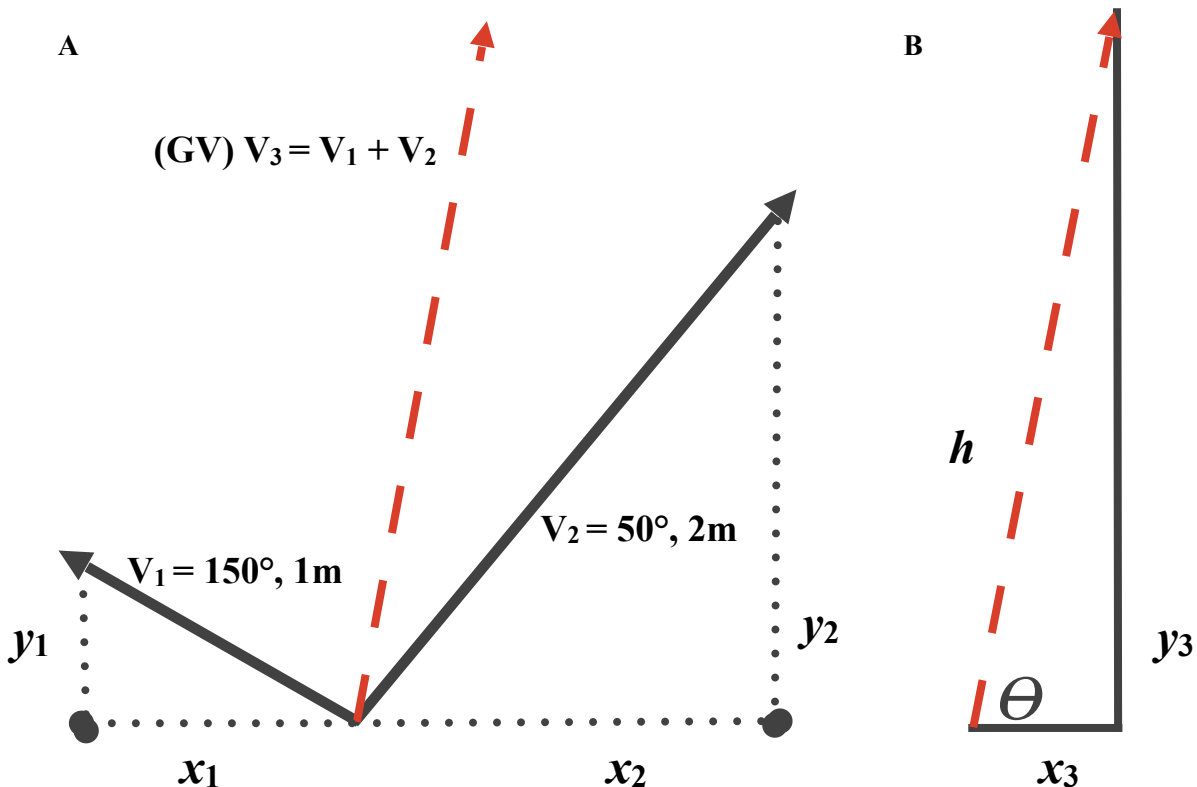
Fig. **26:** An illustration of component analysis of two given vectors (**A**), and the subsequent right angle triangle the summed components complete (**B**) with the global vector, side **h** at angle Θ, still needing to be solved for. The values of the components, and formulae for deriving them, are depicted in *Eq.* **1**.

A

$$x_1 = \cos(150°) \times |V_1| = -0.86 \times 1 = -0.86 \qquad x_2 = \cos(50°) \times |V_2| = 0.64 \times 2 = 1.28$$
$$y_1 = \sin(150°) \times |V_1| = 0.5 \times 1 = 0.5 \qquad y_2 = \sin(50°) \times |V_2| = 0.76 \times 2 = 1.52$$

B

$$V_3 = (x_3, y_3) = (x_1 + x_2, y_1 + y_2) = (-0.86 + 1.28, 0.5 + 1.52) = (0.42, 2.02)$$

> *Eq.* **1:** Formulae for algebraic component analysis on two given vectors (**A**), to derive values for the summed components which compose the global vector needing to be solved for (**B**).

A

$$\theta = \tan^{-1}\left(\frac{y_3}{x_3}\right) = \tan^{-1}\left(\frac{2.02}{4.2}\right) = 78°$$

B

$$h = |V_3| = \sqrt{x_3{}^2 + y_3{}^2} = \sqrt{(0.42)^2 + (2.02)^2} = 2.06$$

> *Eq.* **2:** Formulae for deriving the angle (**A**) and length (**B**) of the hypotenuse, side **h** in *Fig.* **13b** .

What these functions tell you is the extent to which the angles of the vectors under analysis point in the same direction as the angles they are being compared to. In the algebraic case, only two absolute dimensions are used, *X* and *Y*, so all the vectors being analysed are being compared to 0° (cosine) and 90° (sine) respectively.

The first vector points at 150°, so it has a negative value for shared direction with the *X*-axis (as in *Eq.* **1a**). Plainly, it's facing in almost the complete opposite direction. However, it enjoys a positive value for the *Y*-axis, as it's still pointing upwards to some extent. By modulating the magnitude of V_1 (its length) by the amount in which V_1 points in the same direction as 90° (the sine function) and 0° (the cosine function), you learn how much distance V_1 also covers in those directions; you learn its component vectors.

Therefore, by walking 1m at 150°, V_1 effectively walks a total of 0.5m at 90°, and -0.86m at 0°. Mathematically, this sharing of direction is called the *projection*. The cosine gives you the projection of the *X*-component onto the given vector, and the sine the projection of the *Y*-component onto the given vector.

The global vector is now just the unsolved hypotenuse of a right angled triangle (side **h** in *Fig.* **26b**). Thanks to Pythagoras, this is a tractable problem. We know two sides of this triangle already from the algebraic component analysis performed above (*Eq.* **1b**). The adjacent side X_3 is .42m, and the opposite side Y_3 is 2.02m. To solve for *theta* first, which is the angle that the global vector is pointing in, the appropriate trigonometric function is an inverse tangent of the two known sides, depicted in *Eq.* **2a**. The length of the hypotenuse, i.e. the distance of the global vector, is then obviously given by the

Pythagorean maxim, as in *Eq.* **2b**. We have now integrated our path, deriving a global vector, V_3, which points in the direction of 78°, with a length of 2.06m.

5.2 Encoding, Decoding, & Storing Local Vectors in a Neural Network: A Model of Path Integration

With a preferred depiction of analog computation in the foreground (see **2.3**, **4.2**), and a crash course in the mathematics involved with path integration, it is now possible to begin the connectionist response. The Bernardet et. al. (2008) artificial neural network model represents a biologically inspired neural architecture for path integration in Desert Ants. It models the use of proxies for direction and distance information, their integrated encoding and storage, and finally their decoding into a global vector. An overview of its operation is as follows (see model architecture in *Fig.* **27**).

At each moment, the model takes input from the Heading and Velocity units, and integrates them to signal the direction in which the creature is travelling, and how fast it is doing so, in the Gater system. This is achieved through each of the 36 neurons in the Heading system representing a compass direction, and the firing rate of the single Velocity[20] neuron representing speed of travel. This integrated signal represents both direction and speed of travel. It is, therefore, a representation of the ant's velocity.

The Gater's velocity signal will subsequently activate a directionally matched column in the Memory system, comprised of 36 columns, each containing 100 neurons. Each of the 36 neurons in the Gater system is stochastically connected to all 100 of its corresponding neurons in the directionally matched Memory column. In virtue of this probabilistic distribution of connections between Gater and Memory columns, the Gater signal has a 5% chance of activating any one of the 100 cells in the column corresponding to the current direction of travel. Once activated, the cells in the memory columns remain active until the final integration process begins.

Keeping the cells in each Memory column persistently firing enables the system to stably encode information about the path the ant has taken. This is done by exploiting the relationship that exists between velocity and displacement. Information about changes in velocity, from Gater, can be used to form different, but related, vectorial representations about static paths taken. Again, a path is a distance (i.e. a magnitude) travelled in a particular direction. Greater speed of travel (the magnitude of a velocity vector) naturally implies that a greater distance will be covered in that direction. So, as a particular Gater neuron gets activated more frequently and at higher rates, it will activate more neurons in its matching Memory column. This represents a longer distance travelled in that particular direction. The Memory columns thereby utilise a population code to encode distance, i.e. the lengths of paths, and a topographic coding for their directions, divided into 36 bins of direction. The systems downstream can use those 36 individual representations of paths with varying lengths when the walk is over by summing them together.

---

[20] This is somewhat confusing, as velocity is already a vector quantity containing heading *and* speed information. It would more accurately be called the 'Speed' neuron, to convey its proper function in the model as a simple speedometer.
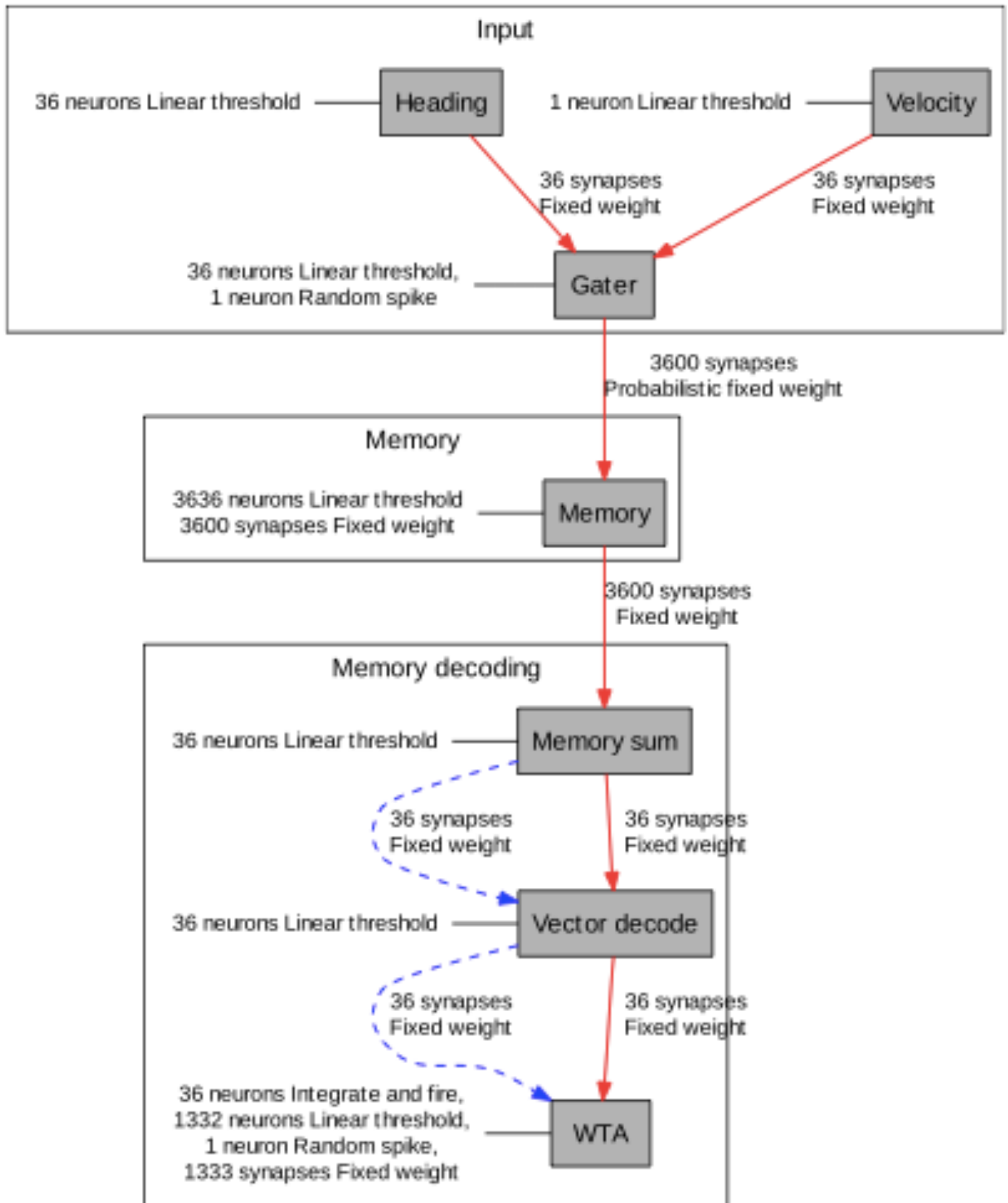
Fig. 27: The Bernardet et. al. (2008: p. 166) neural network model's modular architecture. See text for details.

Summation of the vectors is achieved first by each Memory column activating a corresponding Memory Sum neuron at a rate that is proportional to the number of persistently active neurons the Memory column has. The magnitudinal information about static paths is thereby converted from a population code into a rate code. Each of the 36 neurons in Memory Sum are then connected to all 36 neurons in the Vector Decode layer. However, the synaptic connections these layers are excitatory as well as inhibitory. This is so that a Decode cell is maximally activated by its directionally-matched Sum cell, activated to a lesser extent by less proximal cells, neither inhibited nor activated by cells less proximal again, and finally inhibited by the cells that are farthest away from it.

This process of squeezing the path information in Memory Sum through the particular distribution of inhibitory and excitatory connection weights is what approximates the requisite trigonometric operations. The result is that the most active neuron in the Vector Decode layer represents the global vector, i.e. the total displacement relative to point of departure. Finally, the 'winner-takes-all' (WTA) module determines which of these Vector Decode cells is in fact the most active. This cell, representing a direction and a distance, then merely needs to be inverted to represent the home vector. When tested, the margin of error exhibited by the model was around 20°, which is the margin of error observed in actual ants (Mole, 2014).

5.3 Analogicity & the Structural Representation of Vectors

It's clear why a connectionist would want to use this in reply to Gallistel's worries (**4.1**). It addresses the first problem of computational complexity by only needing 5,187 synaptic connections, in a model that has a mere 5,151 neurons in total (Bernardet et. al., 2008). Clearly, this model demonstrates that connectionists networks need not always be computationally extravagant, like the SMc model. The Bernardet et. al. model is well within the biological range of the Desert Ant. The second, deeper problem concerning information corruption is solved by not utilising reverberating loops at all. It doesn't use a place-code representational strategy like the SMc model to represent and track *absolute* position across a two-dimensional plane. The representational and computational strategy is entirely different; it uses groups of neurons that resemble vector quantities to store information about local outbound paths, and then decodes them to indicate a single path of particular length and particular direction, a global vector.

The first instance where a structural resemblance relation can be seen is in the model's representations of velocity in the Gater system. Vectors have both a direction and a magnitude. Graphically, we often imagine vectors as arrows with various lengths, as this illustratively captures their properties. A longer velocity arrow represents a faster rate of displacement in the direction in which it points. The model, rather than using symbolic encodings, represents these mathematical properties of vectors using systems of neurons that share these same properties.

The integrated signal of the Gater system inherits direction information natively from the initial direction-dependent tuning of the 36 Heading neurons, giving the signal its 'directional' component[21]. The arrow 'lengths' in Gater are then given by their variable activation rates inherited from the speedometer, the Velocity neuron. Having both speed and directional information makes the Gater

---

[21] More on inheriting direction information in sections **5.5 & 5.6**.

signal a physical analog of the mathematical relation it is being used to represent. It is not an 'analog signal', in the sense of being continuous rather than discrete, but an *analog* simpliciter; it possesses the same mathematical properties. Moreover, the relations between the units in Gater preserve the physical relations between the simulated ant's actual changes in velocity. As the activation passes from neuron to neuron, the Gater system tracks shifts in direction topographically, and changing activation rates track changes in travel speed. Crucially, the physical relations that obtain between the ant's actual velocity are preserved in the diachronic patterns of activation realised in the Gater module — an intrinsic, second-order property sharing (see ***Fig. 28***).

Secondly, the Memory columns can be similarly understood as structural representations of paths. The transient information regarding velocity from Gater is used to encode long-term information about paths in the Memory columns. The lengths of these path-vectors are given using a population code though instead of a rate code. As above, the thing in the world needing to be tracked, i.e. the path being traversed, has a direction and a length. The vehicle being used to represent the path has both direction and length information. The physical relations between the units in the Memory columns preserve the relations between paths being taken. As the ant walks farther in a certain direction, the matching column has its active population increased accordingly. The active population of the column gives the length, and the columns are again directionally-matched and arranged topographically, where nearby directions are encoded by nearby columns[22]. One can neatly reimagine ***Fig. 28*** as a depiction of the Memory columns, where the length of the arrows correspond to the proportion of nodes currently active, rather than a single activation rate. Thus, the Memory columns are preserving physical relations between paths walked with their own intrinsic physical properties which correspond to direction and to length.
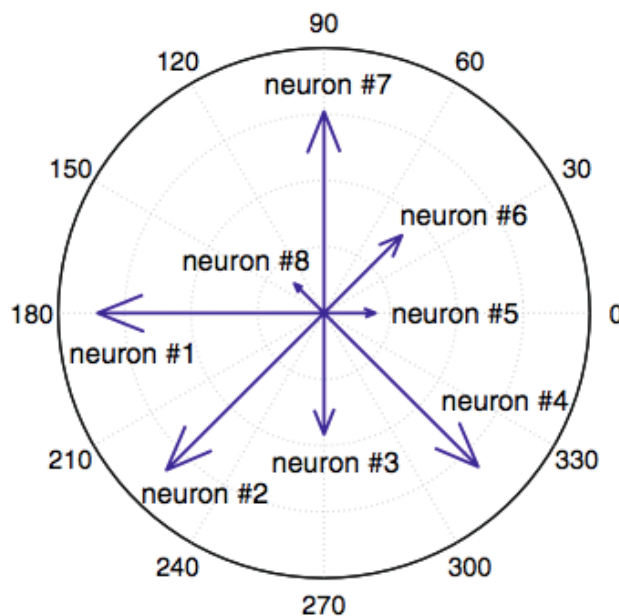

Fig. 28: Illustration of a group of neurons, where each neuron represents a vector that has length proportional to the firing rate and angle given natively by the neuron's direction-dependent tuning, arranged topographically like a compass. (Figure from Bernardet et. al., 2008: p. 165)

---

[22] Again, more on how these nodes inherit direction information in **5.5**.

Lastly, the path vectors in the Memory columns are converted back into a rate code in Memory Sum for computational tractability. The population code is used as a stable, long-term store of path information, but once the walk is done the information needs to be decoded somehow. Rather than computing over the structural representations realised over 3600 neurons in the columns, it can reshape the vectors into a more manageable form over a mere 36 neurons. The Sum layer again keeps the direction information natively, with each Memory column is matched to one specific Sum node, but the length of the vectors is preserved by reverting back to a rate code. The firing rate is proportional to the number of neurons active in each column: if all one-hundred neurons in Memory were active by the end of the outbound task, this would correlate to a maximal firing rate at Sum. The pattern of firing rates across a single layer of 36 neurons is much easier to operate over, but it still preserves the information about actual paths taken by utilising the same kind of structural resemblance relation: each path is again represented by a node with direction and magnitude information.

5.4 Analog Trigonometry & Vector Addition

While possessing structural representations realised as activation patterns is the beginning of an analog account, these are essentially derivative. Possessing groups of neurons that represent velocities (in the Gater module) and paths (in the Memory and Sum modules) is not the same as *computing* over those representations. The model can possess as many structural representations of vectors as it wants, but behaviourally, the all-important computational task is deriving the global vector out of the 36 local vectors cached in its memory.

The structural representations of vectors within this network architecture don't themselves determine how they are to be transformed. They only affect the processing in the sense that they "determine" what is available to be transformed. The maximally active cell in Vector Decode, which putatively is to correspond to the global vector, is a product of computational work. As discussed previously, analog computation requires that the processing itself be governed by, or actively exploit, mathematically comparable laws as those in the task domain. That is, a robust connectionist account requires an intrinsic analogy to exist between the causal dynamics that drive the elements within both systems. What are the underlying *gears* to this model?

The distribution of connection weights between Sum and Decode is what performs this computational labour, and ultimately determines the success of the model. Path integration by vector addition is the crucial lynch pin, so it is here that the connectionist must find analogicity at work. The model's authors write, "[Our] model is biologically plausible in that it does not rely on neurons that perform trigonometric functions. Instead the required trigonometric operations are approximated in the specific connectivity between neurons, and the weights of the connections" (Bernardet et. al., 2008: p. 171). That is, rather than putting in trigonometric activation functions into each neuron, which would be entirely biologically unrealistic, the model puts the trigonometric operations into the overall pattern of connection weights. A graphic depiction of the weights is given in *Fig. 29*. The trigonometric operations are only *approximated* using a "Gaussian" distribution of weights for biological plausibility (Bernardet et. al., 2008). A cosine distribution would be mathematically ideal, but Gaussians are the observed biological norm (Mole, 2014).
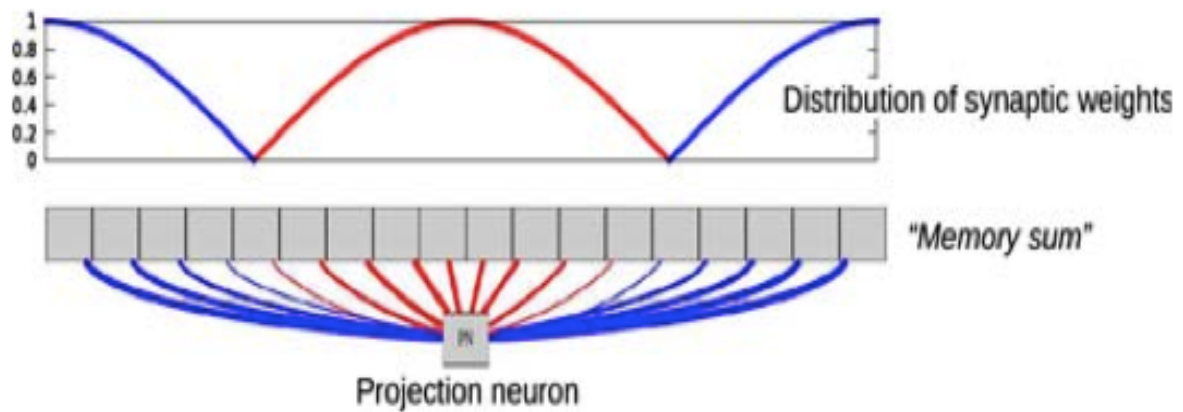
Fig. 29: An illustration of the connection matrix between Memory Sum & Vector Decode. Excitatory connections are in red, inhibitory connections in blue. Top panel shows the weights graphically in their "Gaussian" distribution, which approximates a cosine function. In the bottom panel, strength of connection is represented by width. The "Projection neuron" is one of 36 in Vector Decode. (Figure from Bernardet et. al., 2008: p. 167)

Given that the distribution approximates a cosine function, what it effectively achieves are 36 individual X-component analyses (as in *Eq.* 1a) for each and every neuron in Memory Sum. The magnitude is given by the firing rate, and the cosine function that it is to be modulated by is given, or approximated, by the specific connection weight. Each neuron in Memory Sum essentially takes itself as an *X*-axis. Compare this with the algebraic method in **5.1** where two *absolute* dimensions were used to compare everything else against in a Cartesian plane. So, the connection weight distribution modulates each Memory Sum neuron's activation level 36 times. Each modulation is proportionate to the approximate cosine of the difference in angle between it and the particular neuron in Decode it is connected to. This cosine of their difference in angle is a measure of the degree to which they both point in the same direction, or how directionally similar they are. The Gaussian connection distribution is thus computing the projection of all of the stored local-vectors in Sum onto all of the Decode neurons, which act as basis-vectors. These geometrical relations, concerning orthogonality, or measures of shared directionality and angular sensitivity, are of great significance and will be pervasive throughout the discussion to follow.

The resulting flow of activation through the Gaussian connection matrix, all summed, produces a new set of activation rates in Decode, where one neuron will, under normal circumstances, be more active than all the others. This neuron represents a vector which receives the most modulated activation from all of the neurons in Sum. This maximally-active neuron thus shares the most activation—modulated according to angular similarity—with all of the outbound local-vectors. This is functionally equivalent to the digital component analysis achieved above in section **5.1**. This is a hard concept to express, however. To get a tangible grasp on how this works using the idealised example from *Fig.* 26 again, if the connection matrix embodied an ideal cosine function, rather than an approximate, it would operate as shown in *Fig.* 30.

One can visually observe the 36 individual *X*-component analyses here, and why the weights are distributed as they are. Neuron-15 in *Fig.* 30b receives all of the activation from neuron-15 in *Fig.* 30a, representing the stored data regarding path V_1. These neurons are directionally matched, both are pointing at 150°, so the amount of angular difference between them is 0. Each neuron takes itself as the
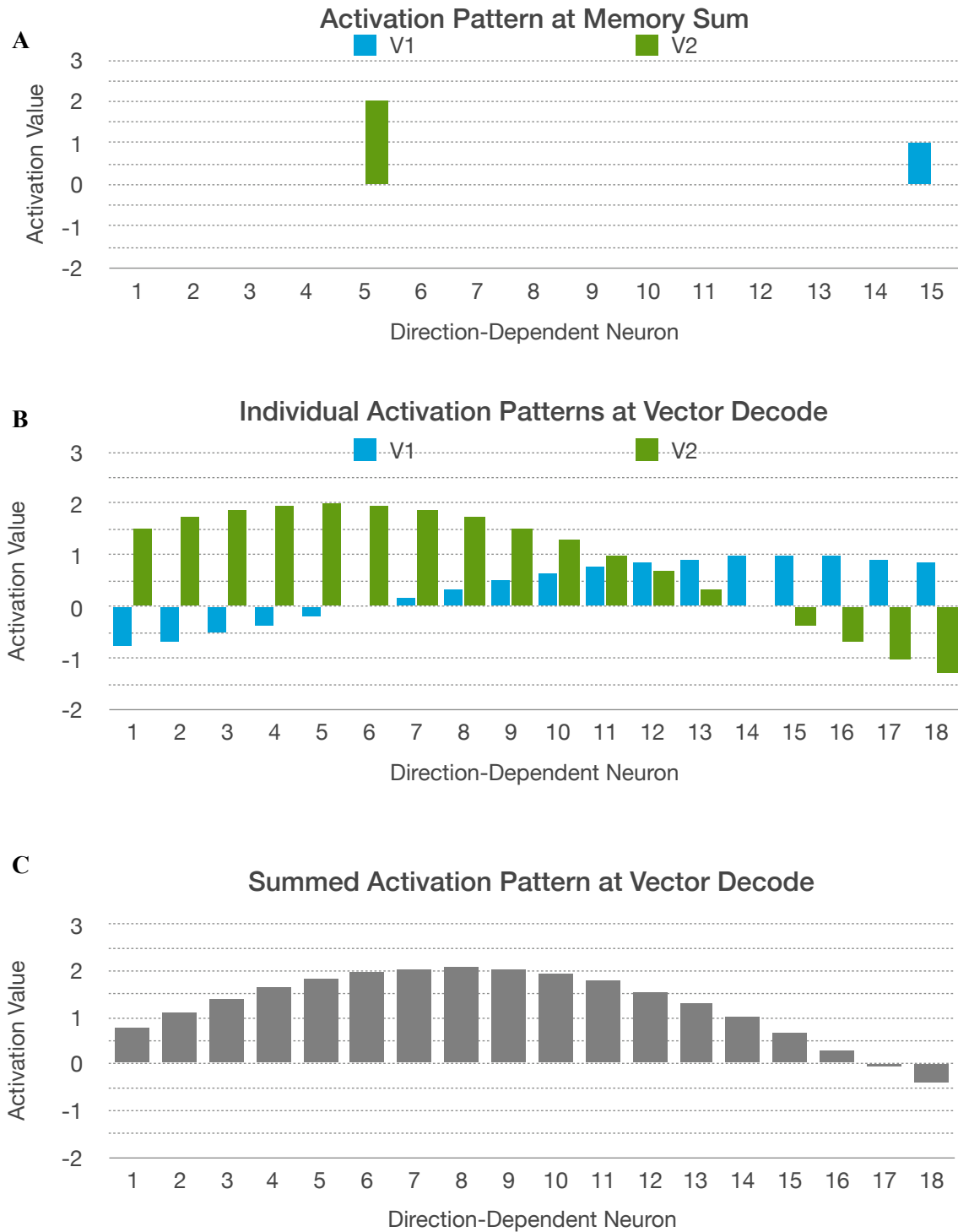
Fig. 30: An illustration of an imagined network with the structure of the Bernardet et. al. model performing vector addition. Using a connection weight matrix that has a perfect cosine distribution, two given vectors (**A**), with length given by activation value and angle by direction-dependent neuron, are modulated (**B**) to derive a summed, global vector (**C**). The depiction in **B** represents the input each Decode neuron would receive individually from the Sum layer (**A**). **C** depicts what the final activation pattern would actually appear as in Decode. Note how the activation levels take on the shape of a cosine wave, a result of spreading out initial activation from only two neurons through the idiosyncratic connection matrix. The two "cosine waves" of activation are then linearly summed at Decode.

"*X*-axis", so the cosine of 0° is 1. Therefore, these two neurons will have a connection weight that modulates activation from Sum (**30a**) to Decode (**30b**) by a value of 1. That is, it won't modulate it at all, instead letting all of the activation pass through freely. As you move one neuron either side of neuron-15 in Decode, you get 10° of angular difference between them and the initial neuron-15 in Sum. The cosine of 10° is 0.98, and so almost all of the activation from neuron-15 in Sum gets transferred to neurons 14 and 16 in Decode, because they almost point in exactly the same direction.

The same neuron-15 in Decode, however, is receiving negative activation from neuron-5 in Sum (***Fig.* 30b**), because neuron-5 represents the vector V_2 that points at 50°. There are 100° of angular difference between these two neurons, and because they take themselves as the *X*-axis, are again connected according to the *cosine* of that angular difference. The cosine of 100° is -0.17, and so the connection between these two neurons modulates the activation value of V_2 in Sum by -0.17, giving a result in Decode of -0.34. The result is negative as the extent to which they point in the same direction is negative. They directionally oppose each other, and the precise extent to which they oppose each other is given by the cosine of the difference in angle between them.

The model of course sums all of these incoming activation values at Decode simultaneously. This summation gives the actual value for each neuron in Decode, as depicted in ***Fig.* 30c**. The most active neuron is neuron-8. This is where most of the activity is accumulating; it is the vector with the most similarity-weighted activation from all the others. It is therefore the most similar vector to the sum of all local vectors taken, i.e. the global vector. It is also firing at a rate of 2.07. This analog example, albeit contrived, is telling us that the global vector, V_3, is pointing at 80° with a length of 2.07m.

Even with a much coarser mathematical and computational grain, this analog approach gets you an answer within 2° and 0.01m of the algebraic method performed in section **5.1**. The model generates the final output because its materials share the relevant fundamental mathematics present in the process it is being used to represent. First, the intrinsic physical properties of the representing vehicles storing path-information preserve a structural resemblance relation with their target domain. Second, the way that the connection weights resemble the required mathematical domain, their varying strengths being distributed like a cosine function, drives the computations involved in deriving the global vector from the local vectors cached in memory (see ***Fig.* 31**).

Of course, the Bernardet et. al. model uses an approximation of a cosine distribution, so the connection weight values, and therefore their computational results, will not be as exact as depicted in ***Fig.* 30 & 31**. Evidently, this doesn't have a significantly deleterious effect on its computational success. The Bernardet et. al. model with its non-ideal, approximated distribution still performs within the observed range of real ants. Critically, we must also remember that the real process being modelled performs this entirely absent of numerical or symbolic representations. Synaptic signs and strengths are real physical/structural features, i.e. physical magnitudes, not symbolic values. An approximation of a cosine function is still deeply analogous to the process it is being used to represent, so it is mathematically comparable and sufficient, not identical or ideal. The relations between the connection weights preserve the relations between an actual cosine distribution, warranting their interpretation as representing vehicles for use by the system in deriving a global vector from its stored cache of local vectors. If this model of the neurobiology involved with path integration withstands further empirical
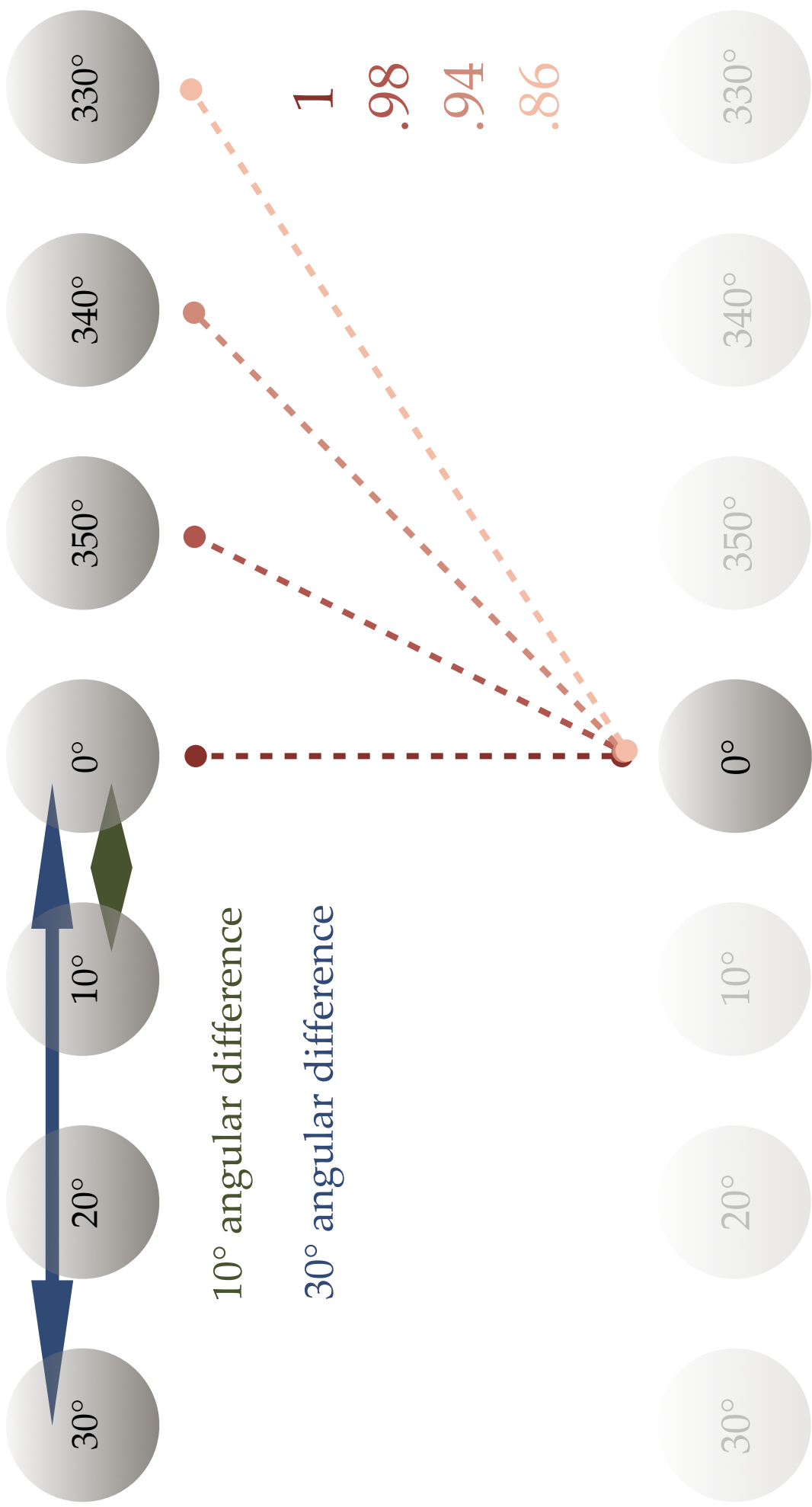
330° 340° 350° 0° 10° 20° 30°

1
.98
.94
.86

10° angular difference

30° angular difference

330° 340° 350° 0° 10° 20° 30°

Fig. **31:** Graphical idealisation of the logic of analog trigonometry in an artificial neural network, based on the architecture of the Bernardet et. al. model. Each directionally-tuned neuron is represent by a grey circle. A single neuron is highlighted, with its connections represented by the red lines of varying strength and opacity, znc their values are given in the key to the right. The rate of change of the connection weights tracks the relationship between the cosines of the angular difference between their respective neurons. That is, their physical structure resembles the requisite mathematical domain in trigonometry. This is an example of 'connection weight representation'.

Fig. 32: Artificial neural network architecture for decoding e-vector signals, presented as triplet codes at input layer, into a population coding of direction of orientation as output, in "compass" neurons. *A*, Representation of network architecture. Activity from the input layer (model POL-neurons) propagates to the output layer in a simple feed-forward antagonistic manner. Black arrowheads represent excitatory connections, white arrowheads inhibitory. Degree values represent the orientational axis of tuning of the respective neurons. *B*, sinusoidal response profiles of the respective layers of model neurons. Dottted points represent the triplet code values taken by the POL-neurons (top) when presented with those particular e-vector orientations. These correspond to the values taken by compass neurons (bottom) at the same orientations. *C*, activity levels of the compass neurons for the same e-vector orientations highlighted in *B*. Compass neuron number corresponds to their arrangement, left to right, in *A*. E.g., compass neuron number nine in *C* (bottom graph) corresponds to the compass neuron for 120° in *A*. (Figures from Sakura, Lambrinos, & Labhart, 2008: p. 669)

They cluster around the approximate angles of preferred orientation in a looser manner than is being modelled here, but this is for purposes of model clarity (Labhart, 1986; 2000). It is nonetheless another non-arbitrary result of the geometrical properties of the DRA receptors, and their specific connection patterns, that real POL-neurons obtain their unique preferred orientations of tuning. The model system at hand removes the unnecessary complexity of the DRA, and sets the POL-neurons' axes of tuning accordingly to account for that absence.

The simulated input, represented to the network as various triplet codes, represents a single artificial e-vector, which is then gradually rotated. We can imagine the triplet as representing magnitudes for the axes of tuning of the POL-neurons, and their combination representing the axis of orientation produced by weighing their various magnitudes together. In other terms, the magnitude for any POL-neuron represents the weighted sum of directional information present in its preferred angle of orientation. Note that the triplet-code input pattern does not contain any canonical information about the azimuthal position of that e-vector in the sky. Rather, this input represents the processed average of direction information filtered through the DRA after sweeping it across the sky. Scaling this up to real world examples, where much more than a single e-vector is available to process, is relatively straight forward. The addition of more e-vectors will just change the activation pattern present at the DRA, and its modulations across time. This will ultimately manifest as an adjustment in the condensed encoding of that activation, i.e. as a different triplet-code at the layer of POL-neurons.

The input POL-neurons then follow simple arithmetic activation and threshold functions, but feed antagonistically into a second layer, which consists of six model neurons. Each of these six neurons receives two connections from the input layer, one inhibitory and one excitatory. However, the tunings of these neurons are non-arbitrary, because the mapping of their connections is non-arbitrary. The tuning axes of the intermediate neurons are created in situ, by selectively combining the *sign* and *rotational axis* of the signals they receive from the input layer.

The result is an expanded set of tuned intermediate neurons, again with sinusoidal activation functions, but with tunings set at 15, 45, 75, 105, 135, and 165°. These intermediate neurons in turn send signals via a mixture of inhibitory and excitatory channels to a third, output layer containing 12 model neurons. Their connections follow the same kind of non-arbitrary mapping of connections as above. This final output layer represents the "compass" neuron layer, where the neurons are now finely tuned to be active when the model creature is facing in a tight wedge of direction, remaining quiescent otherwise. In this case, the compass neurons are 15° apart, spanning 0° - 180°.

The model demonstrates that tunings of 0, 60, and 120° at the input layer can condense and encode what would be a ~180° range of messy e-vector information received by one hemisphere of the visual field, just by altering the amount of magnitude each POL-neuron has and how it projects to subsequent layers. Following from this, it only requires a small amount of neurons, performing simple addition or subtraction with simple connection profiles, to expand that condensed information from a triplet-code back out into a specific heading, using an array of compass neurons that cover 180° of direction information. Of course, in virtue of having a rotational period of 180°, any POL-neuron will therefore fire maximally for two symmetrical angles of orientation. Effectively, this makes them unable to distinguish between a heading of 0° and 180°, or between a heading of 90° and 270°, for example.

However, this problem is solved in natural settings by the scanning behaviour, in conjunction with having two of these systems running in parallel, one for each eye. Each eye's bilateral antagonism between anterior and posterior sections of the DRA (see **3.4**) can kick in and thus affect the output array of compass neurons.

The Sakura, Lambrinos and Labhart (2008) model has the e-vector stimulus being simulated in the same position, but gradually rotating around. A natural setting would have the ant rotate instead, while the stimuli remain static in the sky. The natural setting permits the antagonism between anterior and posterior sections of the DRA to function properly, and consequently makes the POL-neurons that are sensitive to both "antipodal" orientations of polarised light to affect one another. As a result, each eye's DRA and POL-pathway becomes sensitive to only one of those orientations, and a complete compass can then be formed inside more integrative neural structures within the insect's central complex, where the DRA from both hemispheres eventually ramify (Sakura, Lambrinos, & Labhart 2008: p. 673).

The model performance is evaluated by analysing the groups of neurons active at the compass output layer. In *Fig.* **32**, the authors represent the activation profiles for each layer, and output histograms for various orientations of e-vector given as input to the system. As the bottom diagram shows (*Fig.* **32c**), the output activation patterns cluster around a group of neurons. During simulation, the way in which the cluster of active neurons shifts is linearly correlated to the way the e-vector input itself is rotated around (Sakura, Lambrinos, & Labhart, 2008: p. 678). The population median of the active neurons is thus useful as a continuous representation of the e-vector direction given as input. Combined with an initial calibration mechanism for aligning with the anti-solar meridian (see **3.4**), the model as a whole demonstrates how to successfully track, encode, decode, and *exploit* spatial information regarding e-vectors in the sky into a compass or heading cue. Moreover, this model of e-vector coding and translation is biologically realistic, parsimonious, and computationally tractable — and absent symbolic mechanisms.

5.6 Analogicity in the Celestial Compass

The model at hand, and the underlying neurobiology being modelled, is now ripe for connectionist interpretation. This section will argue that the success of the polarisation vision pathway can be reasonably understood as a result of how the neural assemblies involved match*,* track*,* and exploit some of the spatial and geometric features of polarised skylight *in virtue of their own* spatial and geometric features. That is, by way of a resemblance relation.

The first plausible example of analogicity driving this system is the manner in which the microvillar UV-receptors inside each ommatidium are arranged perpendicular to one another. The physical fact of having two sets of receptors arranged like a cross allows each ommatidium to capture an important geometrical relationship; orthogonality. As they both face any given stimulus simultaneously, they become, in effect, biological trigonometricians in virtue of their intrinsic orthogonal alignment. When the e-vector stimulus is perfectly aligned with one of the receptors, that receptor will fire at maximal value, "+1". When the stimulus is orthogonal to one of the receptors, that receptor will fire minimally (i.e. it will be inhibited maximally), "-1". However, by looking at their response curves (see *Fig.* **20**), we can interpret what they're achieving as sending signals that are a continuous measure of "shared

direction" or orientation; a trigonometric projection just like in **5.1**. As the stimulus begins to deviate from the receptor's angle of alignment, the receptor response drops off, i.e. they are signalling that they are facing in the same direction as the stimulus to a lesser and lesser extent, until they reach orthogonality. Orthogonality is the point at which they share *no* directional properties at all.

As in **5.1**, if two vectors being compared have the same directionality, the cosine of 0 (the amount of angular difference between them) is 1, which means they are "completely similar" in their directional properties. Hence, the responses of the neurons involved are maximal, "+1", when both receptor and e-vector stimulus are perfectly aligned. If the two vectors are 45° apart, the cosine of 45 (in degrees) is 0.70. They are therefore "somewhat similar" in directional information. Hence, a reduced response rate in the UV detecting neurons, e.g. "+0.7", respects the precise manner in which the e-vector stimulus' direction deviates away from the receptor's microvillar alignment.

As in *Fig.* **20**, having orthogonal receptors doesn't achieve a perfect mapping between their responses to variously oriented stimuli according to what a real cosine function would dictate. Expecting a neural response profile to perfectly match an abstract mathematical function, however, is probably an unrealistic standard to set. The overall relationship between angular deviation and response strength is still being preserved in the rough-and-ready way exhibited by the receptors though. The orthogonal crossed-receptors are not merely indicating that their stimulus is present, they are giving some graded-response that tracks the extent to which their stimulus is rotationally similar to themselves.

The more profound effect is observed at the shared integrating neuron(s) they project to. The response curve of the integrating 'crossed-analyser' becomes much closer to being sinusoidal, or perhaps more accurately, to being a *cosine* wave[25], due to the interaction of the antagonistic signals it receives. It is in this way that we see a resemblance relation holding between the physical structure of the receptors and the target domain. Linearly polarised light has directionality built into it, hence the term "linearly" polarised light (see *Fig.* **15**). The DRA receptors are sensitive to that directionality of polarised light because of their own inherent directionality; their polarisation-sensitive microvillar arrangement.

The joint functioning of each set of crossed-receptors then allows the integrating crossed-analysers to measure shared direction more sensitively, analysing the orientational match that exists between e-vector stimuli and the angle of orientation of the receptors' microvilli. They thereby perform another biological approximation of a trigonometric operation, a non-symbolic analog of calculating the same kind of result that algebraic component analysis would achieve. One microvillar plane acts as the *X*-axis, the orthogonal counterpart as a *Y*-axis. Each microvilli sends a signal about how similar the e-vector stimulus is to its own axis. Together, their integrated signals amount to computing a "global" vector, the result of summing the two vectorial signals received from the microvilli. Rather than being a measure of net displacement or distance though, as in **5.1**, it is a measure of something like 'net orientational information'. It is a weighted sum that indicates the predominating angle of orientation present in the stimulus. It is in virtue of the orthogonal structure and subsequent integration that the second-order analyser response profiles more closely track a cosine function when presented with various orientations of e-vector stimuli (*Fig.* **20**).

---

[25] The difference being that a sine wave begins at 0 (*sine*(0) = 0), whereas a cosine begins at 1 (*cos*(0) = 1).

The second and perhaps most obvious example of analogicity is the total distribution and arrangement of the polarisation-opponent photoreceptors across the DRA, i.e. the "matched filter". As in *Fig. 21*, they are spread out across the DRA in a fan shape. The rate of rotation that the fan shape has, i.e. the changing angular relation between the detectors, tracks the angular relation between e-vectors in the sky. Each eye then resembles one hemisphere of the averaged celestial pattern of e-vectors. If the DRA detectors were not aligned in this manner (see **3.3**) they would not be able to exploit the spatial information latent in the pattern of polarised skylight in order to navigate. The matched-filter works only because its structure is a physical analog of some of the target domain's own structure.

The rest of the discussion will explore analogicity further "down" the polarisation vision pathway, in layers that the Sakura, Lambrinos and Labhart (2008) network model represents, as described in **5.5**. The model POL input neurons condense and average the information from the DRA and subsequent integrators into three wide-field axes of orientation, "axial vectors" of sorts[26]. As highlighted above, in a natural setting their preferred axes would be tuned in situ, rather than set arbitrarily, in virtue of receiving and thus condensing only particular signs and strengths of signal. I argue that these POL-neurons are inheriting the directional properties from their lower-order counterparts, identical to the way in which the second-order integrators inherit their directionality or tuned axis of orientation from the crossed-receptors at the DRA. The maximum possible response of a POL-neuron, i.e. its preferred axis of tuning, is a direct function of the sections of the DRA which project to it. Each POL-neuron then responds to its stimulus approximately according to a sinusoidal wave, an analog of performing a cosine function. The relation between the responses that all of the POL-neurons exhibit synchronically, in response to variously oriented e-vector stimuli, thereby tracks a geometrical property in the environmental stimulus: they track an averaged sum of the orientational information that exists in the pattern of e-vectors currently present in the sky in each of their preferred axes.

The second and third layers of the network similarly inherit their directional properties as a function of their connection profiles. Consider the neuron farthest to the left in the second layer of *Fig. 32a*. In virtue of having only two connections, an excitatory from the 0°/180° POL neuron and an inhibitory from the 120°/300° POL neuron, this neuron thereby becomes tuned to the 0°/180° axis. Why? The maximum possible response it can exhibit is the maximum possible response of its excitatory predecessor, minus the maximum possible response from its inhibitory predecessor. The two POL predecessors can never be maximally active at the same time, because they're orthogonal to each other. In virtue of the intrinsic angular relation between their two axes of orientation, and the signs of their respective connections, the 0°/180° axis wins out. This process is what is meant when one neuron is said to "inherit" direction information natively from its predecessors, and it applies for the other neurons in this layer as well as neurons in the Bernardet et. al. model in **5.2** (see *Fig. 31*). Their preferred axes of orientation are all determined non-arbitrarily in situ by the particular signs and rotations of signal they receive. This process approximates the appropriate trigonometric function for determining a global vector for net orientation information. The same mathematics of vector addition

---

[26] All that is meant by "axial" vector is that these neurons are not, strictly speaking, representing "vectorial" information, as they are sensitive to an entire axis of orientation. A *vector* has one direction, but that is only a diachronic feature. These neurons are sensitive to *two directions:* a vector and its antipode/inverse. This is functionally appropriate given that a transverse wave of polarised light, the e-*vector*, has both a down- and up-stroke while it propagates forward in its plane of polarisation (*Fig. 15*). The microvilli detectors are obviously not capable of distinguishing between these motions.

applies, because the same underlying geometrical properties are being detected and analysed: direction and magnitude.

The above process is repeated between the second and final layers. The combination of directionally similar neurons sending excitatory signals, and directionally divergent neurons sending inhibitory signals, integrate to shape the preferred axes of activation for the "compass" neurons. The whole feedforward process maintains the sinusoidal response profiles of all the neurons at every layer, as the initial triplet-code is gradually decoded out into select compass neurons at the output layer. This maintains a smooth angular-tracking relationship between the variously tuned neurons' responses to the orientation of the presented stimulus.

The authors note that the second layer is in some sense computationally redundant, but electrophysiological recordings demonstrate that neurons with those profiles do exist in Orthopterans (Sakura, Lambrinos & Labhart, 2008: p. 678). The authors propose that the second and third layers could be combined by merging their connection matrices, but that such a merged matrix would have variable weights, and become "…difficult to comprehend and to express graphically" (Sakura, Lambrinos & Labhart, 2008: p. 678). However, we've already seen what this merged matrix could look like in *Fig.* **30** & **31**, inspired by the Bernardet et. al. (2008) model. Putting in a connection matrix that approximates a cosine function would perform the same computational task: passing vectorial signals forward, modulated according to the cosine of the angular difference between the two neurons involved.

This model deals with essentially the same problem through slightly different means. It still needs to measure the degree of similarity between vectors, but it needs to calculate a global vector that concerns which orientation of e-vector stimulus is predominating over the others, rather than to compute net displacement. In virtue of the underlying mathematics of vector addition being the same, however, using neural structures that share the same kind of vectorial and trigonometric properties operates just as well. The Sakura, Lambrinos & Labhart model just uses simplified connection weights to more gradually decode the relevant vectorial information out. However, the connections' integral role in the overall computational logic is the same: neurons with shared direction receive excitatory input via excitatory connection weights, neurons that are directionally divergent receive inhibitory input via inhibitory connection weights, thus physically embodying the relevant trigonometry underlying the comparison and analysis of vectors. A symbolic read/write mechanism is never appealed to during this processing, nor is it required in order for the neural architecture to compute over its relevant data. We can appeal to, and sensibly understand, how these neural assemblies structurally represent certain geometric-relational aspects of their target domain instead.

Chapter 6
Conclusion
Embracing the Analog Alternative

"Je n'avais pas besoin de cette hypothèse-là."

(*I had no need of that hypothesis.*)

- *Pierre-Simon Laplace*

6.1 Preliminary Conclusions: Matched-Filter vs. Ephemeris-Almanac

The argumentative burden of the prior sections was to demonstrate an alternative to conceiving of Desert Ants (and similar creatures) as navigating by storing strings of symbolic values for variables like "solar azimuth/elevation", and "e-vector azimuth". Such an anthropocentric conceptualisation of the problem domain is biologically unrealistic — *at best*. As far as my survey of the literature indicates, there is no neurobiological process whereby an ephemeris-almanac could be filled-in using such symbolic values. Therefore, it is uncontroversial to reject the hypothesis that they are representing the celestial positions of e-vectors in a stored and symbolically tabulated format. The polarisation vision pathway simply does not record and update real-values for such variables and carry them forward in time through a symbolic memory store, as hypothesised by Gallistel. Indeed, most of the polarisation and other celestial information detected by the creature is destroyed once it passes through the very first layer of sensory apparatus. Evidently, Gallistel's challenges do not set appropriate top-down constraints for how the problem is to be solved, and there are alternative ways to frame the computational logic of path integration. Similarly, the reasons that Gallistel believes connectionism is unable to begin solving this problem, and approach explanations for cognition in general, are therefore entirely unfair. Based on the relevant data, the creatures themselves do not approach the computational task in the manner depicted by Gallistel. So, *why* should a connectionist theorist? Moreover, the connectionist has an alternative framing and interpretation of the problem that is in fact perfectly amenable with the relevant data.

The neural pathways for polarisation vision that were analysed above never match any of the specific and idiosyncratic patterns of polarised skylight to times of day. This is because the relevant sensory apparatus is not actually sensitive to specific idiosyncratic patterns of polarised skylight. It is a "nonimaging" subsystem of the visual modality, where the only kind of information it detects and processes is some *average of e-vector orientations* in the sky (Sakura, Lambrinos, & Labhart, 2008: p. 680). This is a kind of spatial/geometrical *relational* property which is latent within the skylight pattern, but it is not related to the canonical azimuthal *locations* at which any given orientation of e-vector is found in the sky at any given time. Orientations and azimuthal locations of e-vectors are never matched together in this sensory modality.

Instead, polarisation vision in insects like ants involves deploying a generalised template, a fan-like pattern of e-vector detectors, through which polarised skylight is filtered. While this filtering process is insensitive to and indeed *destroys* information regarding the azimuthal positions of specific e-vector directions at any given time, the 'matched-filter' template is sensitive to those latent and navigationally useful geometrical relations that exist in the e-vector patterns. In virtue of the physical relations that obtain between the representing elements of the matched-filter, the polarisation-vision pathway resembles and tracks second-order relational properties in the environment, such as hemispheric (a)symmetries in spatial arrangements, and angular (dis)similarities. The filter is then exploited for local calibration and orientation purposes because it matches those relational features of e-vector patterns in the sky: the internal geometry created by a given e-vector pattern in the sky is approximately matched by the internal geometry of the sensory apparatus. One author summates, "…the e-vector positions assumed by the ant do not coincide with the e-vector positions actually realized

in the sky. From this it is concluded that *Cataglyphis has no detailed knowledge of the actual azimuthal positions of the e-vectors.*" (Fent, 1986: p. 145: emphasis added).

The computational framing that Gallistel has postulated, the stored and "time-matched" pattern hypothesis, also cannot explain experimental results where ants orient and guide themselves when tested with a single, artificially produced e-vector with an orientation that has not occurred in any natural skylight pattern that they have been exposed to (Fent, 1986). Even more reason to suggest that their navigation is independent of stored time-matched pattern "stamps" is that all of the polarisation detection and analysis occurs at the very periphery of the nervous system. The information that survives to more central neural assemblies (where information can be bused to, and used by, motor systems or integrative knowledge-stores, for example) has already been processed and repurposed. The matched-filter immediately changes information received from the e-vector pattern in the sky into bodily-orientation information at the very initial stages of perception, calibrated to the symmetry plane marked by the anti-solar meridian. Thus, after the filtering process, there are no internal representations about the celestial dome and its "veridical" contents. Therefore, the Desert Ant emphatically *can not* be navigating using time-matched idiosyncratic patterns of such contents by registering local correlations between their changes over time. The information regarding any given polarisation pattern's idiosyncratic and canonical distribution is processed and exploited early on in this unique vision pathway, and subsequently filtered out. Wehner concludes that:

> [The] large-field integrators sample the outputs of analyser populations looking at different areas of the celestial hemisphere. The array of compass neurons would then encode particular headings of the ant relative to a celestial system of reference rather than particular e-vectors. Seen in this light, there would be *no stage along the polarization vision pathway at which information about the orientation of individual e-vectors is represented.* (2003: p. 582, emphasis added)

> [The] insect does not…solve the problem by performing a number of abstract computations in some kind of central processing unit. *Its trick is to incorporate the fundamental spatial properties of the navigational problem into the very periphery of its nervous system,* into the spatial design of its sensory surface. (1989b: p. 81-82, emphasis added)

> [In] the cataglyphid ants it was possible to show that the navigational errors occurred only when the animals were trained under a restricted skylight pattern and tested under the full sky, or vice versa. They disappeared completely when the same restricted skylight pattern was present in both the training and the test situation… These behavioural analyses show that desert ants—as well as honey bees—*are not informed exactly about where, within a sun-based system of coordinates, any particular e-vector occurs at any particular time of day.* Nevertheless, their compass system can successfully deal with any partial skylight pattern, as long as the foraging insect can experience the same partial pattern during its entire round-trip. (2003: p. 580, emphasis added)

It is clear that the ethological and neurobiological data on polarisation-vision are at odds with Gallistel's framing of the computational problem to be solved: the local solar ephemeris function—conceptualised literally as a formal function that is inferred from a symbolically tabulated almanac—is not learned or memorised by the Desert Ant. Moreover, the hypothesis that is actually postulated in the relevant empirical literature is entirely amenable to a connectionist interpretation. That is, of course, once connectionism is properly understood as an analog computational framework. A 'matched-filter' is just 'structural resemblance' by another name: it is a sensory apparatus with intrinsic physical-relational properties that resemble some counterpart properties in the problem domain of interest. The polarisation-vision pathway is, I argue, best understood as an example of an analog computational system.

The previous chapters have set out a detailed model of how connectionism and analog computation is, to say the very least, a *viable* framework for understanding the detection and analysis of the polarisation pattern, and the integration of outbound paths for homeward navigation. Sections **5.5** and **5.6** detailed how connectionism would explain the method by which polarisation information is exploited and converted into usable compass cues using the 'matched-filter' and 'scanning hypothesis'. Sections **5.3** and **5.4** demonstrated how a connectionist framework can explain the manner in which those compass cues are integrated with a proxy for speed, in order to store and compute over local vectors. Ultimately, the discussion in **5.4** also showed us how to compute a global vector that represents the path back home — all without using a symbolic read/write mechanism. These models were interpreted as leveraging structural representations of their relevant environmental variables, and demonstrated how these computational tasks can be performed non-symbolically. However, the celestial compass and path integrator models discussed so far are silent on the behavioural evidence showing that Desert Ants appropriately adjust their homeward navigation to account for the long term movements of the sun across the sky.

The second type of restriction experiments (see **3.5**) demonstrated that generalised knowledge about the overall relationship between time and solar-azimuth change clearly isn't learnt, but that some training behaviours do indicate a local learning phase (Wehner & Muller, 1993; Wehner & Lanfranconi, 1981; Grob et. al, 2017). Given the differences in connectionist and classical computation highlighted previously, the process of learning can be conceptualised in a different manner. With a connectionist framework, synaptic weights can be representationally and computationally rich, as introduced in **4.2** and then demonstrated throughout Chapter 5. Connection weights played a robust representational and explanatory role in the processing of the neural network models in Chapter 5. Thus, contra Gallistel, synaptic weight matrices, and therefore any changes therein, can be appropriate theoretical constructs in a computational and representational framework regarding learning and memory, because they can play a pivotal *contentful* role in the formation of some relevant knowledge-structures. This was only possible by moving away from a symbolic representational framework, and using the unique analog computational approach that connectionism can offer. Moreover, given that the ants are not representing the local solar ephemeris function in the manner hypothesised by Gallistel, connectionism can go on in future research to postulate, in its own way, how the ant's innate knowledge about sun-azimuth curves *in general* can be refined via experience to more accurately fit

the *local* sun-azimuth curve. Whatever that mechanism may be, it won't be the kind of mechanism Gallistel has been looking for.

6.2 Broader Conclusions: Modelling Navigation Tasks & the Future

The data from Chapter 3, with the alternative framing of the problem and its solution in Chapter 5, in concert have significant implications for how we can conceptualise and model animal navigation tasks. If taken seriously, these implications can redirect and guide new avenues of research in the topic, as well as cast new light on old problems. First, let's take stock of the alternative framing of the problem of Desert Ant path integration set out so far.

The overall conceptual model of the polarisation-vision pathway that has emerged from this discussion looks something like *Fig.* **33**. The DRA of the Desert Ant compound eye has a fixed template of polarised-light detectors and analysers, distributed in a fan-like shape. These detectors, while insensitive to the azimuthal locations of their stimuli in the sky, are excellent at detecting the average peaks of orientational information present in the overall pattern of stimuli in the sky. Because of this
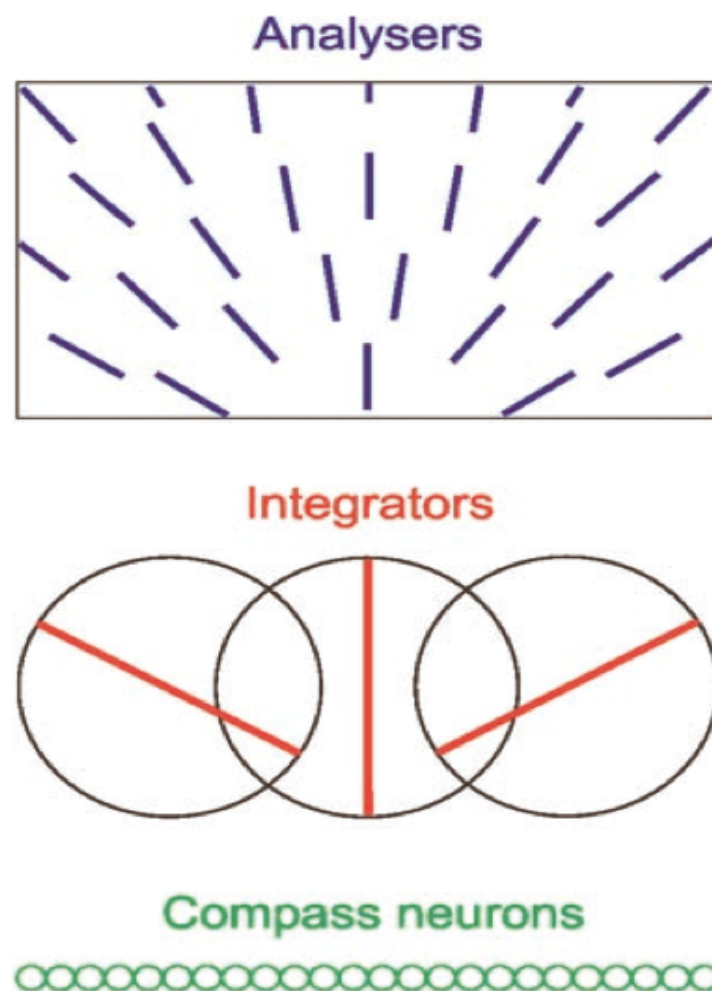

Fig. **33:** Schematic representation of the hypothesised modular, hierarchical structure of the polarisation-vision pathway in the Desert Ant. Top figure represents an idealised set and distribution of polarisation detectors from the initial layer of the DRA in the compound eye of the ant. Middle figure represents the large-field integrators which pool and condense information from specific populations of the DRA analysers. Bottom layer represents the hypothesised compass neurons, each activating maximally only when the creature is facing in a specific angle — with reference to the calibration angle after scanning movements. This visual pathway can then provide orientation or heading information to a subsequent path integrating system. (Figure from Wehner, 2003: p. 582)

capacity to detect the relational geometry latent in their stimuli, the analysers endow the ant with the ability to find and register a local reference angle: the symmetry plane marked by the solar/anti-solar meridian. With a unique reference angle fixed at the beginning of each foraging journey, rotational movements performed by the ant will inform them of their angular deviation from that reference angle. This work is performed by the large-field integrator neurons. By pooling information from specific populations of the DRA detectors, the integrators condense the vast amount of orientation information into just a few dimensions of analysis. With some clever wiring, the condensed information is decoded back out into an array of compass neurons that each have their own unique directional tuning, firing whenever the ant is facing in that specific angle.

This final layer of compass neurons, freshly calibrated for each foraging trip, can now be attached to the path integrating circuit. The 'Heading' module of the Bernardet et. al. model (see ***Fig.* 27**) is exactly the nexus for these two systems. Simply imagine that the 'Heading' module is comprised of the final output layer of compass neurons from the Sakura, Lambrinos and Labhart model (see ***Fig.* 32**). We now have a holistic model of Desert Ant path integration, from the detection and calibration of angles of orientation all the way to the decoding of cached local vectors into a global vector for use in piloting home. All of the modules involved, their neurons and connection matrices, have well substantiated biological precedents, and should satisfy any resource constraints set on real ants due to their computational parsimony. The operations of these modules were ultimately interpreted as using structural representations, the currency of analog computational systems. What does all of this do to our general conceptualisation and logical framing of Desert Ant navigation, and their neurobiological solution to this problem?

Thinking back to the discussion in **2.5**, we can now piece together the philosophical and engineering consequences of this model. The ants determine a locally-unique reference angle at the beginning of each foraging journey, and define all other angles of orientation relative to that point. The use of efferent signals from their legs as a proxy for speed information is a kind of time-based unit of distance measurement. Taken together, the ways in which those pieces of information are being detected and represented suggest one conclusion over others: the ants are more sensibly understood as using a kind of *polar coordinate* system (see ***Fig.* 12**). The compass neurons are used to define the angle, Θ, and the amount of time spent walking in that angle defines the distance, *r*, of each local vector of their outbound journey. As depicted in the Bernardet et. al. model, what accrues in the ant's memory are columns of neurons that represent a selection of angles of orientation with various lengths. The empty memory columns are filled in by transforming *velocity* information into pathing information as they complete an outbound journey. The accumulated vectors representing paths in memory are then added up to determine the global vector indicating the final position relative to the point of departure. This mundane sounding conclusion is actually rather radical, because it goes against the standard wisdom of modelling navigation tasks in animals in several ways.

Cheung and Vickerstaff (2010: p. 5) present a broad unified taxonomy of navigation models, which vary across two dimensions: the coordinate system, and the reference frame (see ***Fig.* 34**). We've already seen the differences between the two standard coordinate systems in ***Fig.* 12**. The reference frame indicates whether the coordinates are centred around the creature itself (ego/auto-centric) or
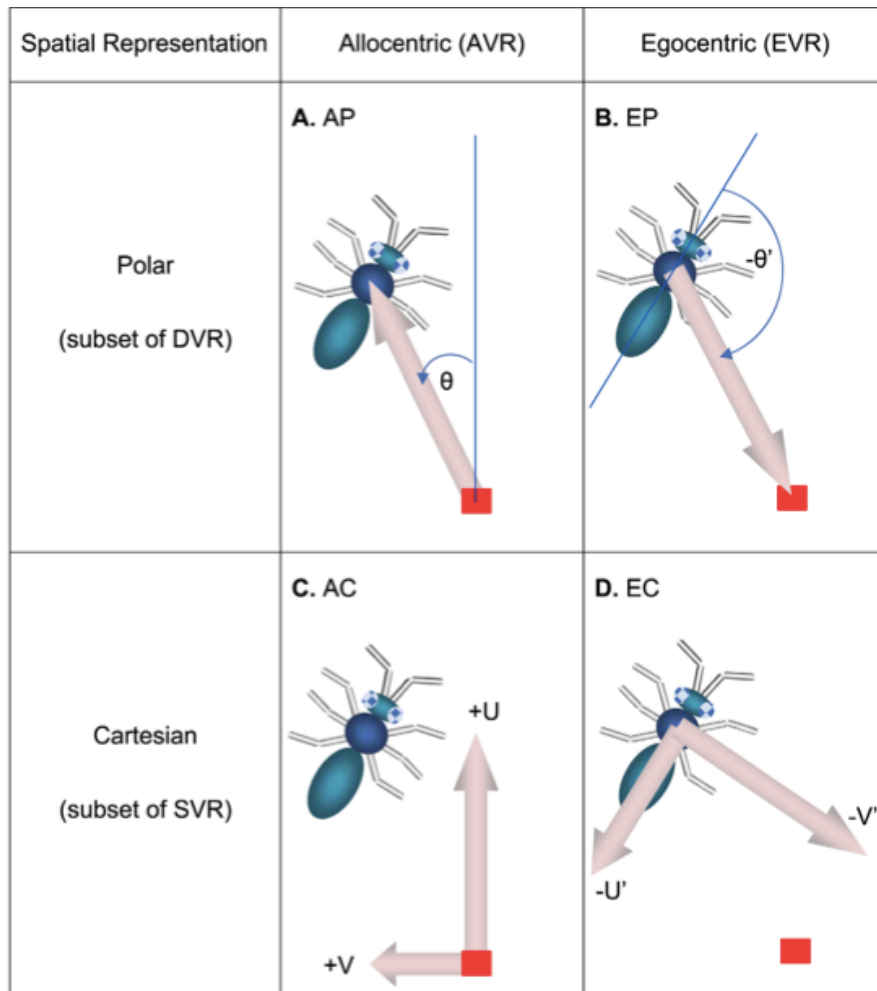
Fig. 34: Taxonomy of the broad classes of models for representing location in 2D space for navigation tasks in animals. The models vary across two dimensions: the type of coordinate system used (polar vs. Cartesian) and where the reference frame is centred (egocentric vs allocentric). DVR: dynamic vector representation. SVR: static vector representation. See text for details. (Figure from Cheung & Vickerstaff, 2010: p. 5)

some external feature (exo/allo/geo-centric). Thus, there are four different ways of representing the same spatial relationship between an animal and its nest. Each has its own advantages and disadvantages regarding variables like resilience to noise and biological realism, and each has its own mathematical assumptions and justifications. The two most important assumptions for our discussion here is the exact manner in which the coordinates are being represented (symbolically vs non-symbolically) and the exact nature of the global or homing vector (HV). It is generally considered that:

> …current theoretical analysis favours a geocentric and Cartesian or static vectorial HV when an allothetic compass cue is available. (Vickerstaff & Cheung, 2010: p. 257)

> …geocentric Cartesian-like systems…appear the most robust solution for implementing a full [path integrating] system (in the sense of modelling the way in which an animal's nervous system needs to process and update information), particularly when an allothetic compass is available. (Vickerstaff & Cheung, 2010: p. 243)

However, such conclusions are drawn because the analysis of the various classes of models were done presuming a *symbolic* medium of representation, and thus a classical mode of computation. That is, formal mathematical equations for the various coordinate systems were analysed for their general

mathematical simplicity and robustness for the task. This is all under the presumption that the neural substrates which have to implement these navigational models in the ant brains would be ultimately performing the same species of computation. If the arithmetic involved gets too unwieldy, produces noise, or otherwise fails to stay within task parameters, then models of neural systems instantiating symbolic functional homomorphisms with such arithmetical operations are dismissed as unviable.

We've seen that this doesn't need to be the case though. Neural assemblies don't have to be interpreted as digital computing systems, and so they don't have to be doing "arithmetic" by pushing symbols around. The neural network models analysed in Chapter 5 lead to an alternative conclusion. The ants are sensibly understood as holding structural representations, not symbolic representations, of their target variables like 'heading' and 'distance'. The problems with the mathematics of polar coordinates must be rethought in light of this fact. The symbolic arithmetic involved with polar coordinates is considered ill-suited for a simple cognitive system, but the structural, geometrical interpretation of performing the same kind of mathematics has never been analysed. The use of connection matrices that physically resemble cosine functions, for example, is a completely new interpretative framework for implementing the mathematics required of coordinate systems and navigation.

Second to this shift in representational media, the nature of the global or homing vector must be reconsidered. As far as I can glean from the literature, the homing vector is without exception considered to be a continually updated, 'live' or 'online' and explicit representation maintained throughout foraging behaviour. That is, as soon as the ant takes a few steps away from the nest, they are thought to have begun integration processes already, and thus have a short, directed homing vector connecting their position to the nest. As they venture further, the integration process continuously updates that vector accordingly. For example:

> Path integration means that the animal is able to *continuously* compute its *present location* from its past trajectory and, as a consequence, to return to the starting point by choosing the direct route rather than retracing its outbound trajectory. (Müller & Wehner, 1988: p. 5287, emphasis added)

> Wherever it goes, the state of its path integrator connects the ant with its home or a place within its foraging area. (Wehner, 2003: p. 583)

> …for directionality and distance information to be behaviourally efficacious, it must be combined so as to calculate the vectors corresponding to each sub-section of the ant's foraging path, and these vectors must be *continually integrated* into a running estimate of the global vector from the ant's nest to its current location. (Morgan, 2014: p. 71, emphasis added)

> [Desert Ants perform path integration] by integrating courses steered and the distances travelled into a *continually updated* home vector. (Wohlgemuth, Ronacher, & Wehner, 2001: p. 795, emphasis added)

The Desert Ant homing vector is demonstrably thought of like a rope, attached to and linking the ant with its home from the very moment it leaves. However, the Bernardet et. al. model anaylsed in

Chapter 5 instead frames the path integration process as a *two-stage* process. First, it accumulates local vectors while foraging. Second, it decodes them once foraging has finished and *only* when the homeward journey is about to begin. The paths accumulated in the Memory columns of the Bernardet et. al. model have implicitly within them a global vector, of course. But it is not until those columns are decoded that the process of vector summation actually begins. If such a model withstands further neurobiological inquiry, this will necessitate a radical change from conventional wisdom: the global, homing vector is not a continuously updated 'live' representation.

Changing the global vector from a continuously updated, online representation to a representation that is only created at a final stage has drastic consequences for the mathematical analyses of navigational models. The details are beyond us here, but a quick example is the 'approach scale' problem. Imagine that the ant is only one metre away from its nest, and then walks a perfect circle around it. Now imagine the ant is fifty metres away from it nest and walks a perfect circle around it. The amount of linear movement that is required to complete the same amount of angular change is different across these two circles, as a mere fact of circular geometry: the fifty-metre orbit around its nest is a larger circle. Thus, theorists have considered that in order to update a polar coordinate system that continually links the creature with its home will have to take into account this fact: as the ant reaches distances farther and farther from home, it will have to down-weight angular change by linear movement. It must then perform the inverse on its way home: as it approaches its nest, the scale increases where smaller units of movement equate to larger changes in angular relation to the nest. This is not only considered computationally complicated in terms of "doing the math" symbolically, but it is an unrealistically burdensome amount of mathematical knowledge to demand that the ant possess.

This problem can be revisited in light of the unique model of navigation given above. The two-stage process has one static homing vector that is used to pilot with, and thus avoids the scale approach problem. It also demonstrates that a polar coordinate system need not be considered a 'dynamic vector' representation. The reference system used by the ant is the external polarisation pattern which is effectively infinitely far away from it, so any movements by the ant across the floor do not change its heading. They are static vectors, which are considered to be the ideal. And yet, this reference system is centred around the creature itself, and calibrated locally for each trip. It is an egocentric reference frame, informed by exocentric cues. I don't pretend to understand the full implications of this (yet), but the standard taxonomy of navigation systems in animals seems to be ill-suited to describe the connectionist model given above. The connectionist approach can thus equip future endeavours with new theoretical tools to revisit and reassess the way we conceptualise navigation tasks in animals.

6.3 Embracing the Analog Alternative

How is information represented and computed over in the brain? Can we conceive of notions like representation and computation in a rigorous, naturalistic way that nonetheless satisfies the founding assumptions of cognitive science? Answering these questions has been the underlying motivation of the present discussion.

As things stand, there are at least two ways to make rigorous and precise the notions of representation and computation. However, one is given short shrift more often than the other. Connectionism can be and is best understood as an independent computational framework, wholly distinct from its competitor, classicism. Connectionism exploits non-symbolic representational structures, and thus does not constitute a digital implementation of a computational device — the kind we're most familiar with, and the kind that connectionism often gets conflated with. The independent connectionist framework posits the use of *structural* representations instead, and thus an *analog* model of computation (see O'Brien & Opie, 2004; 2005; 2008).

When connectionism is understood in this way, it is possible to grant one of the central criticisms levelled against connectionism: malleable synaptic conductances are incapable of instantiating symbolic representational data structures as part of a classical read/write mechanism. However, such criticisms are entirely beside the point: connectionism does not claim to instantiate classical symbolic mechanisms. Instead, synaptic weights can and should be understood as structural representations, like the activation pattern representations that are standard-fare in connectionist theory. That is, synaptic strengths can be elements in a represent*ing* set of physical magnitudes, wherein some physical relations *between* these elements track, and can be mapped onto, some physical relations between elements in the environment, a represent*ed* set. Groups of connections, with varying strengths, can instantiate structural homomorphisms with elements of their task domain in the same way that activation patterns are commonly considered to do so. Structural homomorphism is the content determining relationship that underpins structural representations. The structural homomorphism between sets of elements is useful and productive to some user of the represent*ing* set of elements, insofar as the user is brought into some contact with a represent*ed* set of elements in the world as a result of leveraging that structural homomorphism. So, when Gallistel criticises connectionism, and asks of it, "where are the symbols?", the connectionist simply replies: "I have no need of that hypothesis". There is more to representation than symbols and functions.

The structural and analog framework for understanding connectionism was set out in Chapters 2 and 4, but the cash value of this framework was demonstrated throughout Chapter 5. Desert Ant navigation, and animal navigation tasks more generally, is an old problem. While the relevant sciences of animal navigation have made significant developments over the decades, there have remained gaps in our holistic understanding of the task. The connectionist approach threw new light on these issues.

For some theorists, for example Gallistel, there is a fundamental tension between the mechanisms that have been discovered by the relevant sciences that underwrite Desert Ant navigation, and how we piece those engineering puzzle-pieces together with our philosophical understanding of computation and intelligence. One small change completely reopened this dialogue, to clear (I hope) the tension: there is another way to understand biological computation.

A close look in Chapter 3 at the geophysics, geometry, and astronomy of celestial navigation cues revealed that there are useful yet unusual ways to leverage the contents of the sky for orientation: the relational structure of polarised skylight, rather than their explicit patterns, possess some stable and some asymmetric features with which to calibrate one's orientation. With the proper connectionist

framework of computation in hand, it was then possible to see how some of that relational structure produced by environmental stimuli can be mirrored in the neurobiology of an evolved creature.

The artificial neural network models analysed in Chapter 5 are where the rubber meets the road for the connectionist approach. By looking at some simplified but realistic models, inspired by the neurobiology analysed in Chapter 3, the physical differences were made clear between the connectionist and classical approach. There are plenty of theoretical differences to be drawn out between the two approaches, but I hope to have shown that by holding them up against real biology, and real tasks, the engineering consequences that differ between them really become apparent.

The neurobiology of path integration looks nothing like a classical mechanism. Theorists like Gallistel, as well as the ant scientist-in-chief, Rudiger Wehner, would seem to agree with me on this point. Gallistel simply pursues this conclusion to a different end. He asserts if that we can't find the classical mechanisms he believes are ineluctable constraints on cognitive phenomena with the theoretical tools at hand, tools like patterns of activation or synapse strengths, then we need new tools. Wehner, on the other hand, marvels at how evolution stumbled upon a 'matched-filter' to complete the task of path integration. Rather than using classical mechanisms which perform, "…a number of abstract computations in some kind of central processing unit" (Wehner, 1989b: p. 81-82), the matched-filter incorporates the structure of the task domain into the structure of the sensory apparatus itself. I take Wehner's conclusion, and simply take it one step further in order to integrate it with the philosophical, computational foundations of cognitive science: a 'matched-filter' is a quintessential analog, *connectionist* computational mechanism. Achieving conciliation between the empirical data and a philosophically robust interpretation then gives us a well-rounded cognitive explanation of Desert Ant navigation.

It was possible to make sense of the information encoding strategy in the ant's neurobiology using structural representations. We thereby gained insight into the ant's perspective — the way the ant takes the world as being. More precisely, it became possible to make sense of the rational way the ant behaves given the way it represents its own relation to the world. Connectionism can, therefore, illuminate reason-based interpretations required of cognitive, intelligent systems. At the very least, it can explain *basic* cognitive phenomena like path integration in insects, contra the claims made by Gallistel.

Scaling these efforts up, to include total explanations of brain functioning (more than just highly circumscribed navigation mechanisms using only a few modules) will be a long slog for the connectionist programme. I hope to have taken some small steps in the present discussion to show, at the very least, that such a programme is *possible*. My more extravagant hope is that I've shown how such a programme is theoretically *promising*. Connectionism is a viable research programme in the cognitive sciences, and the way that it unpacks notions like representation and computation are radically different to the standard conceptions. These radical differences in turn have consequences for the way we think about cognition, and can reshape and revitalise the way we investigate and model cognitive systems. To follow through on this promise, and see where the connectionist framework can take us, we need to relieve ourselves of a parochial focus on symbolic, digital computation, and embrace the analog alternative.

References

Azhar, F. (2016). Polytopes as vehicles of informational content in feedforward neural networks. Philosophical Psychology 29 (5):697-716.

Bartels, A. (2006). Defending the structural concept of representation. Theoria 21 (55):7-19.

Bechtel, W., & Abrahamsen, A. (2002). Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks. Wiley-Blackwell.

Beck, J. (2014). Analogue Magnitudes, the Generality Constraint, and Nonconceptual Thought. Mind 123 (492):1155-1165.

Bernardet, U., Bermúdez, I., Badia, S., & Verschure, P. F. (2008). A model for the neuronal substrate of dead reckoning and memory in arthropods: a comparative computational and behavioral study. Theory in Biosciences 127(2): 163–175.

Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. Synthese 193 (5): 1509-1534.

Campbell, G. (2015). Memory & The Computational Brain with Randy Gallistel. [podcast] Brain Science Podcast. Available at: https://www.youtube.com/watch?v=zj6nutVdN6c.

Chalmers, D. (1995). Facing up to the Problem of Consciousness. In Journal of Consciousness Studies 2: 200-19.

Cheung, A., & Vickerstaff, R. (2010). Finding the Way with a Noisy Brain. PLoS Computational Biology, 6(11).

Churchland, Paul M. (1998). Conceptual similarity across sensory and neural diversity: The fodor/lepore challenge answered. Journal of Philosophy 95 (1):5-32.

Churchland, P., & Sejnowski, T. (1990). Neural representation and neural computation. Philosophical Perspectives, 4, 343–382.

Clark, A. (2015). Predicting peace: The end of the representation wars—A reply to Michael Madary. In T. Metzinger & J. M. Windt (Eds.), Open MIND: 7(R). Frankfurt am Main: MIND Group.

Dacke, M., Nordström., P., & Scholtz C. (2003). Twilight orientation to polarised light in the crepuscular dung beetle Scarbaeus zambesianus. Journal of Experimental Biology, 260:1535–1543

Davidson, D. (1975). Thought and talk. In S. Guttenplan (Ed.) Mind and Language, (p. 7–23). Oxford: Oxford University Press.

Dennett, D. (1987). True believers. In The Intentional Stance, (p. 13–36). Cambridge, MA: MIT Press.

Dretske, F. (1988). Explaining Behavior: Reasons in a World of Causes. Cambridge, MA: MIT Press.

Duelli, P., & Wehner, R. (1972). The Spectral Sensitivity of Polarized Light Orientation in Cataglyphis bicolor (Formicidae, Hymenoptera). Journal of Comparative Physiology, 86: 37-53.

Elman, J., et al. (1996). Rethinking Innateness: A Connectionist Perspective on Development. Cambridge, MA: MIT Press.

Fent, K. (1986). Polarized skylight orientation in the desert ant Cataglyphis. Journal of Comparative Physiology A, 158, 145-150

Fodor, J. (1975). The Language of Thought. Harvester Press, London

Fodor, J. (1987). Psychosemantics: The Problem of Meaning in the Philosophy of Mind. MIT Press.

Fodor, J. (1992). The big idea: can there be a science of the mind? Times Literary Supplement July 3: 5–7

Fodor, J. (2008). LOT 2: The Language of Thought Revisited. Oxford University Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. Cognition, 28, 3–71.

Gallistel, C. (1998). Symbolic processes in the brain: The case of insect navigation. An invitation to cognitive science, 4, 1–51.

Gallistel, C. (2008). Learning and representation. In J. Byrne (Ed.) Learning and Memory: A Comprehensive Reference, vol. 1, (pp. 227–242). Amsterdam: Elsevier.

Gallistel, C., & King, A. (2010). Memory and the Computational Brain. Oxford: Wiley-Blackwell.

Gallistel, C., & Matzel, L. (2013). The neuroscience of learning: Beyond the hebbian synapse. Annual Review of Psychology, 64(7).

Grob, R., Fleischmann, P., Grübel, K., Wehner, R., & Rössler, W. (2017). The Role of Celestial Compass Information in Cataglyphis Ants during Learning Walks and for Neuroplasticity in the Central Complex and Mushroom Bodies. Front. Behav. Neurosci. 11:226.

Hardwick, C. (Ed.) (1977). Semiotics and significs: The correspondence between Charles S. Peirce and Victoria Lady Welby. Bloomington, IN: Indiana University Press.

Hebb, D. (1949). The Organization of Behavior: A Neuropsychological Theory. New York, NY: Wiley.

Horgan, T. (1997). Connectionism and the philosophical foundations of cognitive science. Metaphilosophy 28 (1-2), 1-30.

Horgan, T., & Tienson, J. (1992). Cognitive systems as dynamic systems. Topoi 11 (1), 27-43.

Isaac, A. (2013). Objective Similarity and Mental Representation. Australasian Journal of Philosophy, 91/4: 683 704.

el Jundi, B., Pfeiffer, K., & Homberg, U. (2011). A Distinct Layer of the Medulla Integrates Sky Compass Signals in the Brain of an Insect. PLoS ONE 6(11).

Labhart, T. (1986). The electrophysiology of photoreceptors in different eye regions of the desert ant, Cataglyphis bicolor. Journal of Comparative Physiology A, 158:1-7.

Labhart, T. (1988). Polarization-opponent interneurons in the insect visual system. Nature 331:435–437.

Labhart, T. (2000). Polarization-Sensitive Interneurons in the Optic Lobe of the Desert Ant Cataglyphis bicolor. Naturwissenschaften, 87, 133–136.

Labhart, T., & Meyer, E. P. (1999). Detectors for polarized skylight in insects: a survey of ommatidial specializations in the dorsal rim area of the compound eye. Microsc Res Tech 47:368–379.

Labhart, T., Petzold, J., & Helbling, H. (2001). Spatial Integration in Polarization-Sensitive Interneurones of Crickets: A Survey of Evidence, Mechanisms and Benefits. Journal of Experimental Biology, 204, 2423–2430.

Maley, C. (2017). Toward Analog Neural Computation. Minds & Machines.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: W.H. Freeman.

McClelland, J., & Rumelhart, D. (eds) (1986) Parallel distributed processing: explorations in the microstructure of cognition, Vol. 2. MIT 108Press, Cambridge.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115–33.

Mole, C. (2014). Dead Reckoning in the Desert Ant: A Defence of Connectionist Models. Review of Philosophy & Psychology, 5:277–290.

Morgan, A. (2014). On the Matter of Memory: Neural Computation and the Mechanisms of Intentional Agency. PhD Dissertation, Rutgers University.

Müller, M., & Wehner, R. (1988). Path integration in desert ants, Cataglyphis fortis. Proceedings of the National Academy of Sciences Jul 1988, 85 (14), 5287-5290.

Nestor, M. J. (2016): Narrow Structuralism: Paving a Middle Path Between Cummins and Millikan. Australasian Journal of Philosophy.

O'Brien, G. (2015). How Does Mind Matter? Solving the Content Causation Problem. In T. Metzinger & J. M. Windt (Eds). Open MIND: 28(T). Frankfurt am Main: MIND Group

O'Brien, G., & Opie, J. (2004). Notes towards a structuralist theory of mental representation. In P. S. Clapin & P. Slezak (Eds.) Representation in mind: New approaches to mental representation. Amsterdam, NL: El- sevier.

O'Brien, G., & Opie, J. (2005). How do connectionist networks compute? Cognitive Processing 7 (1): 30-41.

O'Brien, G., & Opie, J. (2008). The role of representation in computation. Cognitive Processing 10 (1): 53-62.

Piantadosi, S. T., & Jacobs, R. (2016). Four problems solved by the probabilistic Language of Thought. Current Directions in Psychological Science, 25, 54–59.

Rossel, S., & Wehner, R. (1982). The bee's map of the e-vector pattern in the sky (polarized light/astronavigation). Proceedings of the National Academy of Sciences, 79: 4451-4455.

Rossel, S., & Wehner, R. (1984). How bees analyze the polarization patterns in the sky: Experiments and model. Journal of Comparative Physiology A, 154: 607-615.

Rossel, S., & Wehner, R. (1986). Polarization Vision in Bees. Nature, 323(11), 128-131.

Rumelhart, D., & McClelland, J. (eds) (1986) Parallel distributed processing: explorations in the microstructure of cognition, vol. 1. MIT Press, Cambridge.

Sakura, M., Lambrinos, D., & Labhart, T. (2008). Polarized Skylight Navigation in Insects: Model and Electrophysiology of e-Vector Coding by Neurons in the Central Complex. Journal of Neurophysiology, 99: 667-682.

Samsonovich, A., & McNaughton, B. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. Journal of Neuroscience, 17(15), 5900–5920.

Thomson, E., & Piccinini, G. (2018). Neural Representations Observed. Minds and Machines 28 (1): 191-235.

Vickerstaff, R., & Cheung, A. (2010). Which coordinate system for modelling path integration? Journal of Theoretical Biology, 263, 242–261.

Von Eckardt, B. (1993). What is cognitive science? Cambridge, MA: MIT Press.

Wehner, R. (1987). Spatial organization of foraging behavior in individually searching desert ants, Cataglyphis (Sahara Desert) and Ocymyrmex (Namib Desert). In: JM Pasteels, J-L Deneubourg (eds) From individual to collective behavior in social insects. Birkhäuser, Basel, pp 15–42

Wehner, R. (1989a). Neurobiology of Polarization Vision. Trends in Neuroscience, 12, 353 -359.

Wehner, R. (1989b). The Hymenopteran Skylight Compass: Matched Filtering and Parallel Coding. Journal of Experimental Biology, 146, 63-85.

Wehner, R. (2003). Desert ant navigation: how miniature brains solve complex tasks. Journal of Comparative Physiology, 189, 579 -588.

Wehner, R., & Lanfranconi, B. (1981). What do ants know about the rotation of the sky? Nature, 293(29).

Wehner, R., & Müller, M. (1993). How do ants acquire their celestial ephemeris function? Naturwissenschaften, 80(7), 331–333.

Wehner, R., & Rossel, S. (1985) The bee's celestial compass—a case study in behavioural neurobiology. Fortschr Zool, 31:11–53.

Wehner, R., & Srinivasan, M. (2003). Path integration in insects. In K. Jeffrey (Ed.) The Neurobiology of Spatial Behaviour, 9–30. Oxford University Press.

Wehner, R., & Wehner, S. (1986). Path integration in desert ants: Approaching a long-standing puzzle in insect navigation. Monitore Zoologico Italiano, 20(3), 309– 331.

Wehner, R., & Wehner, S. (1990). Insect Navigation: Use of Maps Or Ariadne's Thread? Ethology, Ecology, & Evolution (2), 27-48.

Whishaw, I. & Brooks, B. (1999). Calibrating space: Exploration is important for allothetic and idiothetic navigation. Hippocampus, 9(6), 659-667.

Williams, D. (2017). Predictive processing and the representation wars. Minds and Machines.

Williams, D., & Colling, L. (2017). From symbols to icons: the return of resemblance in the cognitive neuroscience revolution. Synthese 195 (5):1941-1967.

Wittlinger, M., Wehner, R., & Wolf, H. (2006). The ant odometer: Stepping on stilts and stumps. Science, 312(5782), 1965–1967.

Wohlgemuth, S., Ronacher, B., & Wehner, R. (2001). Ant odometry in the third dimension. Nature, 411(14).