The logic of evolution and the evolution of logic

Integrating the Extended Evolutionary Synthesis and later-Wittgensteinian semantics in an evolutionary model that explains conceptual change

Gijs Anton van der Gun
4163885
March 11, 2022

Faculty of Science
Department of Mathematics
MSc History and Philosophy of Science

Supervisor: dr. Abigail Nieves Delgado
Second supervisor: dr. Erik Stei

Abstract
In this thesis, I develop an evolutionary model that aims to explain conceptual change. Attempts at doing so have generally been called *evolutionary epistemologies*. These focus on constructing analogies between the way species evolve, and the way knowledge changes. I claim this approach has its limits, and I abandon it in favour of one that grounds conceptual change in the evolution of organisms. It can be seen as an extension of evolutionary theory. This extension is twofold: I first use a framework called the Extended Evolutionary Synthesis to ground what I call *practices* in the evolution of organisms, and I then use the later Ludwig Wittgenstein's ideas on meaning and language-games to ground concepts in practices. The result is a model that explains the interdependencies between the evolutions of three notions: *organisms*, *practices*, and *concepts*. In the last two chapters, I apply the model to change in knowledge of logic, and of the notion of logical consequence in particular. In this application, I focus on the logical theories of Aristotle and Chrysippus, and claim that these are grounded in the practices of *giving counterexamples*, and of the *contradiction*, respectively.

Contents

Introduction

When asked to defend a theory, we are likely to point at the available evidence. If two theories are about the same evidence – in other words, when they compete – one theory can be said to be better than the other if the one theory explains the evidence better than the other.[1] This can be a reason to abandon old theories for new ones. Another motive to change theories is in the light of *new* evidence. This is one way to describe change in knowledge: we either come up with new explanations, or we discover more evidence.

The role of evidence is not always so clear for logical theories. There are those who hold that logic is *a priori* – that our justification of logical theories does not rely on experience.[2] Others claim that logic is *a posteriori*, that our logical beliefs *are* justified by experience.[3] This raises the question – not one necessarily unique to logic – of what exactly should count as evidence for competing logical theories. If logical theories cannot compete at all, then this causes issues for the normativity of logic. Logic is not about how we reason, but how we *should* reason. Why should I reason according to *this* logical theory rather than another?

The specific issue that concerns me in this thesis is how to explain change in knowledge – and in the final chapters, change in knowledge of logic in particular. Inferences that were once held true by logicians in the past, are now considered to be invalid. Why is this so? One explanation relies on us humans as fallible epistemic subjects. Assuming that ancient logical theories can indeed be said to compete with modern ones in some way, the logicians in the past were *wrong* – not necessarily through any fault of their own – and we are now *right*, or at least *less wrong*.

Rather than starting my research from this tempting starting point, I instead begin by considering an existing theory that explains genesis and change in another area of inquiry: evolutionary theory. Scholars since Charles Darwin have seen similarities between change of knowledge and change of species, and have generally put forward their ideas on this under the label *evolutionary epistemology*. In this thesis, I propose an evolutionary model to explain conceptual change. My model aims to answer the question of why it is that our concepts of the world around us have changed – or have not changed – so much in the history of knowledge production.

While most attempts at an evolutionary epistemology have focused on applying analogies between change in species and change in knowledge, my model should instead be seen as an *extension* of evolutionary theory. This extension is twofold. First, using a framework called the Extended Evolutionary Synthesis, I introduce the idea of *practices* to evolutionary theory: intentional, replicable behaviour of agents. Second, using the later Ludwig Wittgenstein's theory of meaning, I ground *concepts* – which I assume to be the basic entities which knowledge is made up of – in practices. The result is a model that is based on three notions – of *organism*, *agency*, and *concept* – whose evolutions are interdependent.

---

[1] For an account of this type of 'inference to the best explanation', see e.g. Timothy Williamson, 'Abductive Philosophy', *The Philosophical Forum* 47, no. 3–4 (2016): 263–80, https://doi.org/10.1111/phil.12122.

[2] See e.g. Paul Boghossian, 'Knowledge of Logic', in *New Essays on the a Priori*, ed. Paul Boghossian and Christopher Peacocke (Oxford, UNITED KINGDOM: Oxford University Press, Incorporated, 2001), 229–54, http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3052772.

[3] See e.g. Ole Thomassen Hjortland, 'Anti-Exceptionalism about Logic', *Philosophical Studies* 174, no. 3 (2017): 631–58, https://doi.org/10.1007/s11098-016-0701-8.

In chapter 2, I first discuss earlier attempts at an evolutionary epistemology, which are generally *adaptationist* in that they do not allow much room for agency. In section 2.2, I discuss two *non-adaptationist* approaches to evolutionary epistemology, by Franz M. Wuketits and David L. Hull. In section 2.3, I use their ideas on bidirectional causation and organisms as interactors to introduce the framework of the Extended Evolutionary Synthesis, which enables me to grant organisms the agency necessary to be able to talk of intentional practices.

I start chapter 3 by addressing criticism by Joseph Fracchia and Richard Lewontin on the limitations of creating models by analogy. By using Wittgenstein and the Extended Evolutionary Synthesis to extend evolutionary theory, I claim to avoid most of Fracchia's and Lewontin's worries. In section 3.2, I explain which Wittgensteinian tools I require for my model. This gives rise to the idea of *pre-theoretic notions* – starting points to base our theories and conceptualizations on. In section 3.3, I discuss as an example the pre-theoretic notion of colour, and its categorization into discrete terms. I end the chapter with section 3.4, where I explain what it means to take my model seriously, and to apply its own concepts to itself.

In chapter 4, I explain the model. I start with expounding on the three notions of *organism*, *agency*, and *concept*, and how they give rise to three distinctions: between the organic and the inorganic, between passive attributes and active practices, and between demonstrations and concepts. Using these three distinctions, the model explains the interdependencies of the evolutions of organisms, practices, and concepts. I end chapter 4 with some remarks on how to consider the issue of *correctness* in light of my model – an issue I alluded to in the beginning of this introduction.

Chapter 5 is the start of an application of my model to logical theories and concepts. In section 5.1, I introduce logic as a subject. In section 5.2, I discuss some different notions of *logical consequence*. In section 5.3 and 5.4, I discuss the views of respectively Stewart Shapiro and Graham Priest on the legitimacy of logical theories, and use their views to point out potential misinterpretations of my model.

In Chapter 6, I apply my model to two logical theories in ancient history: Aristotle's syllogistic logic, and Chrysippus' Stoic logic. For both, I follow the same structure. I first identify the relevant practices that I claim gave rise to their respective logical theories. Then, I show that the practice existed before they came up with their logical theories. Third, I provide arguments that make plausible that it was indeed this practice that gave rise to their logical theory – which is *not* to say that they were aware of this. Finally, I link the concept of logical consequence to some other related concepts of *proof*, *truth*, *entities* of truth, and *meaning*.

In the conclusory chapters, I provide a suggestion for a further application of my model to Gottlob Frege's logical theory. I end the thesis with an overview of some of the limitations of my model.

Chapter 2: The Extended Evolutionary Synthesis and its place in the Evolutionary Epistemology

Chapter 2 situates my thesis in the debate on evolutionary epistemology of the second half of the 20th century. It is specifically meant to show that in evolutionary epistemology, evolution is often talked about from an 'adaptationist' perspective, and to show how thinking of evolution in the terms of the Extended Evolutionary Synthesis might provide an evolutionary epistemology different from these existing attempts. Keeping this in mind, I provide a limited overview of the debate on what Michael Bradie calls "the evolution of theories program (EET)" as opposed to the "evolution of cognitive mechanisms program (EEM)."[4] The EET program concerns itself with whether evolutionary theory can explain the development of knowledge, and is not to be confused with the EEM program, which concerns itself with whether evolutionary theory can explain the development of the cognitive mechanisms that give rise to this knowledge. I use Bradie's paper "Assessing Evolutionary Epistemology" as a starting point, in which he discusses attempts at creating an EET program (by Konrad Lorenz, Donald T. Campbell, Karl Popper, and Stephen Toulmin, among others).[5]

In section 2.1, I discuss the adaptationist perspective in evolutionary epistemology. In section 2.2, I consider two evolutionary epistemologists' non-adaptationist perspective. In section 2.3, I explain what the framework of the Extended Evolutionary Synthesis can provide evolutionary epistemology.

2.1 Adaptationist Evolutionary Epistemology

In this section, I argue that most of the attempts in the EET program have followed what can be called an 'adaptationist' approach. I use Michael Bradie's 1986 paper "Assessing Evolutionary Epistemology" to make this point, but see Nathalie Gontier's 2006 *Evolutionary Epistemology, Language and Culture* for a more recent book that in its introduction provides a similar argument.[6] In the preface, she urges to take newer evolutionary theories seriously in the quest for an evolutionary epistemology, and distinguishes between traditional evolutionary epistemology and new evolutionary epistemology – claiming the former takes "an adaptationist point of departure, and hence emphasize the active role of the environment, while the latter subscribe[s] to an organismic point of view."[7] In section 2.3, I expound on this organismic point of view, and in chapter 4, I develop my model accordingly. Gontier's book includes papers that argue for this organismic point of view in evolutionary epistemology, but these do not develop a model that explains conceptual change, like mine does.

According to Michael Bradie's overview of Evolutionary Epistemology, most evolutionary models of conceptual change have three components: variation, selection, and transmission/retention.[8] This is

---

[4] Michael Bradie, 'Assessing Evolutionary Epistemology', *Biology and Philosophy* 1, no. 4 (1986): 403.

[5] Bradie, 413–34.

[6] Nathalie Gontier, Jean Paul Van Bendegem, and Diederik Aerts, eds., *Evolutionary Epistemology, Language and Culture: A Non-Adaptationist, Systems Theoretical Approach* (Dordrecht: Springer, 2006), https://link.springer.com/book/10.1007/1-4020-3395-8. For Gontier, the main difference between what she calls 'traditional' and 'new' evolutionary epistemology, is that the former focuses on the active role of the environment, and the latter on the active role of the organism.

[7] Gontier, Van Bendegem, and Aerts, x–xi.

[8] Bradie, 'Assessing Evolutionary Epistemology', 417.

unsurprising, as these are precisely the three components you can find comprising a "recipe for change" in the form of natural selection.[9] One such recipe comes from R. Levins and R. C. Lewontin:

> 1. There is variation in morphological, physiological, and behavioral traits among members of a species (the principle of variation).
> 2. The variation is in part heritable, so that individuals resemble their relations more than they resemble unrelated individuals and, in particular, offspring resemble their parents (the principle of heredity).
> 3. Different variants leave different numbers of offspring either in immediate or remote generations (the principle of differential fitness).[10]

The question left for evolutionary epistemologists to answer is to what degree this recipe can be extended to non-biological things like concepts. The method to do so has often been the construction of an analogy between evolution of organisms and evolution of science, not only for *variation*, *reproduction*, and *fitness*, but for a number of key biological concepts, e.g. 'organism,' 'population,' 'phenotype,' 'genotype,' etc.[11] However, Bradie points out that most evolutionary models do not provide much detail beyond analogies for variation, transmission/retention, and selection (roughly corresponding to steps 1, 2, and 3 Levins' and Lewontin's recipe, respectively).[12]

Note that Levins' and Lewontin's recipe in itself makes no reference to adaptations.[13] Nevertheless, the assumption has often been that organisms themselves play no active role in the recipe: they are merely entities being optimized by the process of natural selection, and all their traits are adaptations as a result of this process. And like these traits are fitted to the environment, so too are theories fitted to their environment – in whatever way this environment is defined. Karl Popper claimed that "theories become *better adapted through natural selection*: they give us better and better information about reality."[14] Donald T. Campbell supported Popper's view that scientific knowledge develops through an eliminative process analogous to natural selection.[15] Konrad Lorenz thought that the scientific method was akin to the "method of the genome, perpetually making experiments, matching their results against reality, and *retaining what is fittest*" and differed only in that "man also learns from his failures."[16] Stephen Toulmin thought that science consists of competing theories whose survival is *determined by a*

---

[9] Peter Godfrey-Smith, *Darwinian Populations and Natural Selection* (Oxford: Oxford University Press, 2013), 1.
[10] Richard Levins and Richard Lewontin, eds., 'Adaptation', in *The Dialectical Biologist* (Harvard University Press, 1985), 76; quoted in Godfrey-Smith, *Darwinian Populations and Natural Selection*, 18.
[11] Bradie, 'Assessing Evolutionary Epistemology', 416–17.
[12] Bradie, 417. See page 418-421 for a table of these analogies according to different authors.
[13] In fact, together with Stephen Jay Gould, Lewontin famously criticised the adaptationist school. See S. J. Gould et al., 'The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme', *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205, no. 1161 (21 September 1979): 581–98, https://doi.org/10.1098/rspb.1979.0086.
[14] Karl Raimund Popper, 'Evolutionary Epistemology', in *Evolutionary Theory: Paths into the Future*, ed. Jeffrey W. Pollard (Chichester West Sussex: Wiley, 1984), 239; quoted in Bradie, 'Assessing Evolutionary Epistemology', 407. Emphasis mine.
[15] Bradie, 'Assessing Evolutionary Epistemology', 407.
[16] Konrad Lorenz, *Behind the Mirror: A Search for a Natural History of Human Knowledge.* (London: Methuen, 1977), 24; quoted in Bradie, 'Assessing Evolutionary Epistemology', 406–7. Emphasis mine.

selection process.[17] In these examples, the main role is reserved for the selection process, and the organisms or theories are entities passively being selected.

That this adaptationist focus poses the same problem in biological change as in conceptual change was known to Campbell.[18] Nevertheless, the idea of a passive organism being molded by its environment remains prevalent in this adaptationist stance. The organism is thought to be passive in the sense that the variation it provides compared to other organisms of the same species or population, is *random* or *blind* – not only in its development but also in its actions.[19] Whatever environment the organism lives in, is the arbiter of this organism's success. In that way, William James claims that variation in knowledge is that "which the outer environment simply confirms or refutes, adopts or rejects, preserves or destroys – *selects*, in short, just as it selects morphological and social variations due to molecular accidents of an analogous sort."[20] Some authors, like Campbell, Popper, and Plotkin, even turn the analogy on its head, claiming that biological adaptations are a form of knowledge.[21]

Lewontin summarises this adaptationist view as one in which the environment is "causally prior to, and ontologically independent of organisms," and in which "[t]he world is divided into causes and effects, the external and the internal, environments and the organisms they 'contain'."[22] This focus on environmental fit is not a necessary consequence of constructing an analogy based on Levins' and Lewontin's recipe. But Lewontin's critique is not that a theory of biological evolution should not be (metaphorically) applied to scientific change, but that both biological evolution and scientific change are described by the same, erroneous metaphor of trial-and-error adaptations, in which the organisms (or genes) provide the trials, and environmental pressures determine success and failure.[23] Trials in biology do not always lead to increased fitness, and certainly not always to maximized fitness. Natural selection is not that powerful a process, and the environment is not that influential.

None of this is to say that there have not been authors who have noted that natural selection, or its conceptual analogy, should not be expected to explain *everything*; nor that there is not some sense in which both biological evolution and scientific change *are* trial-and-error processes. Nevertheless, the assumption is that, like our knowledge, organisms are a 'true' (but simplified/imperfect) representation of the world: true in the sense that their traits reflect the environment they – through trial and error – developed them in. Otherwise, they would not have survived – otherwise, our knowledge would not have survived. By starting off the EET program with this adaptationist, trial-and-error metaphor, the program has been pushed into a corner it does not need to be in.

I now discuss two authors who are explicitly non-adaptationist in some sense or another: Franz M. Wuketits and David L. Hull. Like me, Wuketits notes that "[e]arlier versions of evolutionary epistemology

[17] Bradie, 'Assessing Evolutionary Epistemology', 408.
[18] Bradie, 417.
[19] Bradie, 422–24.
[20] William James, 'Great Men, Great Thoughts, and the Environment', *Atlantic Monthly* XLVI, no. CCLXXVI (1880): 456; quoted in Bradie, 'Assessing Evolutionary Epistemology', 424.
[21] Bradie, 'Assessing Evolutionary Epistemology', 430–31.
[22] Richard Charles Lewontin, 'Organism and Environment', in *Learning, Development, and Culture: Essays in Evolutionary Epistemology*, ed. H. C. Plotkin (New York: John Wiley & Sons, 1982), 159; quoted in Bradie, 'Assessing Evolutionary Epistemology', 434.
[23] Bradie, 'Assessing Evolutionary Epistemology', 432–33.

were based on — or at least strongly informed by — the adaptationist paradigm."[24] Both strip the environment of its "causal priority" and "ontological independence."[25] I do not address Wuketits' and Hull's full views on evolutionary epistemology, but rather use some of their concepts – organisms as interactors, and bidirectional cause and effect – as a segue to my own model.

2.2 Non-adaptationist Evolutionary Epistemology

Franz M. Wuketits' paper "Evolutionary epistemology: the non-adaptationist approach" is one explicit attempt at providing a non-adaptationist account of EET.[26] Wuketits touches on some aspects I introduce in my model in chapter 4. I also consider David L. Hull's conception of organisms as *interactors* as opposed to *vehicles* (for genes).[27] Hull, like Wuketits, shifts the focus from environment to organism. These two moves – putting the organism front and centre, and recognizing bidirectional cause and effect – are two important facets of my model. They are in line with the Extended Evolutionary Synthesis, which I discuss in section 2.3.

Wuketits remarks that organisms are clearly *active*, rather than mere *passive* objects formed by selective pressures from the environment.[28] Organisms are active by constraining their own parts, as well as by effecting change in their environment. Wuketits calls it a "constant flux of cause and effect in two directions."[29] The point is that causation does not only flow from environment to organism. There is also (bidirectional) causation within the multiple levels that the organism is made up of, and more importantly, causation from the organism back to the environment.

That is not to say that the whole concept of adaptations is useless. Wuketits is careful to observe that a non-adaptationist approach is not necessarily an anti-adaptationist approach.[30] The way an organism has evolved, *is* connected to the environment it evolved in. The main takeaway is that causation does not only flow from environment to organism, but also from organism to environment.

This relates strongly to Hull's conception of *interactors*. In his work "Interactors versus Vehicles," Hull goes against a conception of evolution that features *replication* as the main process in selection.[31] Of course replication *is* important – see step 2 in Levins' and Lewontin's recipe – but for selection to work Hull claims a second process is necessary: *interaction*.[32] Hull defines interactors as "those entities that interact as cohesive wholes with their environments in such a way as to make replication differential."[33] This opposes the view of organisms as mere *vehicles* for the entities being replicated (genes) – or in Wuketits' terms, it gives organisms an *active* role in the process of selection.

---

[24] Franz M Wuketits, 'Evolutionary Epistemology: The Non-Adaptationist Approach', in *Evolutionary Epistemology, Language and Culture: A Non-Adaptationist, Systems Theoretical Approach*, ed. Nathalie Gontier, Jean Paul Van Bendegem, and Diederik Aerts (Dordrecht: Springer, 2006), 33, https://doi.org/10.1007/1-4020-3395-8.
[25] See my earlier comments about Lewontin's summary of the adaptationist approach.
[26] Wuketits, 'Evolutionary Epistemology: The Non-Adaptationist Approach'.
[27] David L. Hull, 'Chapter 2: Interactors versus Vehicles', in *The Role of Behavior in Evolution*, ed. Henry C. Plotkin (Cambridge, Mass: The MIT Press, 1988), 19–50, https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=48504&site=ehost-live.
[28] Wuketits, 'Evolutionary Epistemology: The Non-Adaptationist Approach', 37.
[29] Wuketits, 38.
[30] Wuketits, 40–41.
[31] Hull, 'Chapter 2: Interactors versus Vehicles'.
[32] Hull, 27.
[33] Hull, 27.

Hull thinks that if replication and interaction both play an important role in biological evolution, they do so too in conceptual evolution.[34] He names testing as the analogue for interaction, and thereby also puts the person doing the testing in focus: the scientist.[35] I agree with Hull that interaction *is* important, and so are the practices performed by the scientist. But Hull conflates the two and sees the interaction as something being done by the scientist. In my model, both the practices of the scientist and the idea of organism-environment interaction play an important role, but they are not the same thing. I get back to that in chapter 4.

2.3 What does the Extended Evolutionary Synthesis have to offer the EET program?

In this section, I use Wuketits' notion of bidirectional flux of causation and Hull's conception of interactors to introduce the Extended Evolutionary Synthesis (EES), a view of evolution alternative to the adaptationist stance I outlined in section 2.1. I show that the way EES differs from the adaptationist stance is in line with Wuketits' and Hull's non-adaptationist ideas as I outlined them in section 2.2. The goal is not to argue that the EES provides a better view of evolution than the adaptationist stance, but to show how it provides a *different* view that could inform the EET program in a different way.

According to Laland et al., one of the two key themes of the EES is *reciprocal causation* (the other theme being constructive development).[36] Reciprocal causation is the idea that organisms have an active role to play in evolution, as opposed to being passively selected due to environmental pressures. Causation flows not only from environment to organism, but also from organism to environment.

As I explained in section 2.1, the adaptationist stance is that all causal significance is attributed to the environment. As an example, consider how the adaptationist stance views a phenomenon like niche construction, which is "the process whereby the metabolism, activities and choices of organisms modify or stabilize environmental states, and thereby affect selection acting on themselves and other species."[37] That this happens is not denied by anyone – just look at the beaver building its dam. The difference between the adaptationist stance and the EES lies in how they attribute causality.

According to the adaptationist stance, the building of a dam – thought undoubtedly affecting the beaver's survival – should not be seen as an *evolutionary cause* of change to the beaver population.[38] Instead, the dam can be explained as an 'extended phenotype': just like the polar bear's white coat has evolved because it enhanced fitness, so too evolved the beaver's dam-building because it enhanced fitness.[39] The polar bear and the beaver both play a passive role in the evolution of their phenotypes.

According to the EES, the beaver – through its dam-building – has an active, causal role to play in its evolution. Niche construction "directs evolution by non-random modification of selective environments."[40] In section 2.2, we have seen this bidirectional causal link from organism to environment and back in Wuketits' work. The beaver is not just a passive object formed by selective

---

[34] Hull, 41.
[35] Hull, 44.
[36] Kevin N. Laland et al., 'The Extended Evolutionary Synthesis: Its Structure, Assumptions and Predictions', *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1813 (2015): 6–7, https://doi.org/10.1098/rspb.2015.1019.
[37] Laland et al., 4.
[38] Laland et al., 4.
[39] Laland et al., 5.
[40] Laland et al., 5.

pressures from the environment, but actively changes said environment. In Hull's terms: the beaver is an *interactor*, and its dam-building as interaction with its environment is as necessary for selection to do its work as (genetic) replication is.

So, the EES differs from the adaptationist stance in that for the former, interactions between organisms can actually *drive* the selection process. There is disagreement not on which processes exist in nature, but which of those processes should feature as *causes* when explaining evolutionary processes.[41] Laland et al., using terms borrowed from Peter Godfrey-Smith, compare the adaptationist stance to the EES as follows: the former is 'externalist' in that adaptations of organisms are described in terms of the *external* environment, while the latter is more 'interactionist', in the sense that "organismal (and, for that matter, environmental) change is described […] relative to an organism, rather than to a pre-established environment."[42] It is not simply the environment that should be considered, but the environment in light of the interactions that the organism has with it.

Laland et al. point out that in many cases, niche construction phenomena have not been proven to affect the evolution of the constructor.[43] I think the interesting takeaway is not that a niche construction activity directly influences the evolutionary trajectory of an organism's lineage, but that the environment is the result of countless activities and choices of all organisms in that environment. All these interactions form a "network of causation and feedback."[44] In section 2.1, I quoted Lewontin as saying that the adaptationist stance sees the environment as "causally prior to, and ontologically independent of organisms."[45] If the EES is correct, there is no reason to attach such importance to the environment. What is the environment, if not a collection of other organisms that are relevant for one organism through the interactions it has with them? What other way for the environment is there to have influence on the evolutionary trajectory of an organism's lineage, except for direct, bidirectional, *causal* interaction with it?

This idea of evolution as entailing a network of causation and feedback can also be found in Baedke's et al.'s paper "Unknotting reciprocal causation between organism and environment."[46] In this paper, they attempt to visually model the causal feedback between organism and environment. They propose an 'open-loop' model with two series of states, one for the organism, one for the relevant environment, and arrows between these states that stand for causal processes.[47] Activity of an organism at state O_{n-1} (causally) influences the relevant environment at state E_n, for example by including new organisms that compete with the organism (at state O_n). This then has an effect on the organism's survival,

---

[41] Jan Baedke, Alejandro Fábregas-Tejeda, and Francisco Vergara-Silva, 'Does the Extended Evolutionary Synthesis Entail Extended Explanatory Power?', *Biology and Philosophy* 35, no. 1 (January 2020): 20, https://doi.org/10.1007/s10539-020-9736-5.
[42] Kevin N. Laland et al., 'More on How and Why: Cause and Effect in Biology Revisited', *Biology and Philosophy* 28, no. 5 (September 2013): 730, 737, https://doi.org/10.1007/s10539-012-9335-1.
[43] Kevin Laland, Blake Matthews, and Marcus W. Feldman, 'An Introduction to Niche Construction Theory', *Evolutionary Ecology* 30, no. 2 (April 2016): 194, https://doi.org/10.1007/s10682-016-9821-z.
[44] Laland, Matthews, and Feldman, 195.
[45] Lewontin, 'Organism and Environment', 434.
[46] Jan Baedke, Alejandro Fábregas-Tejeda, and Guido I. Prieto, 'Unknotting Reciprocal Causation between Organism and Environment', *Biology & Philosophy* 36, no. 5 (October 2021): 1–29, https://doi.org/10.1007/s10539-021-09815-0.
[47] Baedke, Fábregas-Tejeda, and Prieto, 14.

development, or reproduction at state O_{n+1} (see figure 1). Note that the states can represent the same organism in time, but also a lineage of organisms.

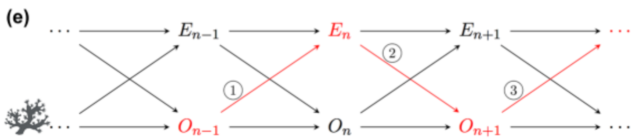

Figure 1. An example of a simple 'open-loop' model involving niche-construction by corals. Corals at state O_{n-1} emit certain chemical compounds, which change the species composition in the environment at state E_n (1). The new species compete with the corals, having effect on the coral state at O_{n+1} (2), and so on (3). (Baedke, Fábregas-Tejeda, and Prieto, fig. 2e).[48]

I think the idea of reciprocal causation and feedback is on the right track to capture the ideas of Wuketits and Hull I introduced in section 2.2. If it does not make sense to talk of the One (causally prior and ontologically independent) Environment, then the notion of organisms as a "true" (but simplified/imperfect) representation of their environment – see section 2.1 – is not as convincing as it seemed prima facie. What is left of the adaptationist trial-and-error metaphor, then? In nature, trials of organisms are contextual: the organisms determine their environment through their actions. In other words, trials consist of the interactions of organisms with their environment, which is defined by these interactions in two ways: the environment is that part of the network that is causally relevant in light of the interactions, and the interactions change the environment.[49] Errors are those interactions that are not reproduced, in spite of the active role that the organism plays in these interactions. So, by choosing and changing the causally relevant part of the network, organisms, through trials, change what errors are.

An analogy with conceptual change could still work, but it needs an account of what it means for the definition of errors to change based on the trials. How do concepts interact with the reality we presuppose them to represent? If this reality is supposed to select for the concepts that are correct, then what does it mean for the concepts and their interaction to change the definition of correctness – of a failed or successful trial? The attempts at an EET program we have seen in section 2.1 do not come this far. They just assume that there is something like theories being "better adapted," as Popper put it.[50]

That does not mean all is lost for the EET program. "[T]he notion of correspondence has to be replaced by the notion of coherence," says Wuketits.[51] We just need a different way to relate evolution to conceptual change. I introduce the necessary tools for this in chapter 3, taken from the later Ludwig Wittgenstein's ideas on meaning.

First, I end this chapter by addressing a possible objection: even if the idea of reciprocal causation and feedback is sufficient to model the evolution of, say, coral, could it also model the complex behaviours of humans? Are those complex behaviours the result of evolution? What exactly *is* the difference

---

[48] Baedke, Fábregas-Tejeda, and Prieto, 15.
[49] Emphasizing the "mutual interpenetration of organism and environment," Levins identifies six ways in which organism and environment are related. See Richard Levins, 'Coexistence in a Variable Environment', *The American Naturalist* 114, no. 6 (December 1979): 766, https://doi.org/10.1086/283527.
[50] Popper, 'Evolutionary Epistemology', 239; quoted in Bradie, 'Assessing Evolutionary Epistemology', 407. See section 2.1.
[51] Wuketits, 'Evolutionary Epistemology: The Non-Adaptationist Approach', 39.

between the activities humans do and those that other organisms like corals do? If we should call one form of change in organisms 'evolution' but not the other, on what basis should we do so? The idea is that there is some form of *cultural* change in organisms that is distinct from biological evolution. There might be important differences between different kinds of change in organisms, but I believe this distinction is not so easily made, as I explain below.

For one, humans can gain new behaviours not only by being born, through (genetic) inheritance, but also throughout their lives, by learning. But the capability to learn is not unique to humans, and a network model like Baedke et al.'s is equipped to deal with such interactions that affect organisms within one lifetime. What about *social* learning? Other animals display social learning too, though perhaps to a lesser extent than humans do. However, the distinction between individual and social learning is not as clear-cut as it might seem. Tim Lewens points out that what we learn individually is influenced by others, at the very least by their past actions.[52] Especially for humans, a lot of the environment we interact with has been built by others. Social learning, however defined, is merely one of the "numerous ways in which activities of one generation can, by altering or maintaining stable features of biotic, social and technical environments, have an influence over what individuals in the following generations end up learning," says Lewens.[53] According to Laland and O'Brien, niche construction – a phenomenon I discussed in section 2.3 – can capture the idea of cultural inheritance too.[54] The distinctions between different ways in which environments are altered or maintained are blurred, and a network model does not have to make these distinctions, as long as there is a causal influence from one organism to the next.

Note that the influence altered or maintained environments have, does not have to flow from one generation to the next. Clearly, if I build a house for my family, that has influence on their lives and activities *now* as well as later. Could the relevant distinction be the one between vertical (from generation to generation) and horizontal (within one generation) transmission, where biological evolution would only encompass vertical transmission, and cultural evolution also horizontal transmission? Well, even genes can transfer horizontally, which already complicates matters and lessens the usefulness of the distinction.[55] Furthermore, niche construction activities seem to transfer both horizontally and vertically: they help maintain a beneficial environment for (a population of) organisms, but they can also help create environments for the development of offspring.[56]

For the purposes of this thesis, I therefore assume a network model to be able to capture different ways – that Lewens has called biotic, social, or technical – in which stability in the network is maintained or altered. Whether this is too big of an assumption remains to be seen, but I think the assumption is sufficiently backed by the Extended Evolutionary Synthesis as proposed by Laland et al.. In section 2.3, we have seen that in the network model of Baedke et al. (see figure 1), it does not matter whether state

---

[52] Tim Lewens, 'Cultural Evolution', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2020 (Metaphysics Research Lab, Stanford University, 2020), sec. 13, https://plato.stanford.edu/archives/sum2020/entries/evolution-cultural/.
[53] Lewens, sec. 13.
[54] See Kevin N. Laland and Michael J. O'Brien, 'Cultural Niche Construction: An Introduction', *Biological Theory* 6, no. 3 (September 2011): 191–202, https://doi.org/10.1007/s13752-012-0026-6.
[55] See for a definition of horizontal gene transfer: Kara Rogers, 'Horizontal Gene Transfer', in *Encyclopedia Britannica*, 22 August 2019, https://www.britannica.com/science/horizontal-gene-transfer.
[56] Laland, Matthews, and Feldman, 'An Introduction to Niche Construction Theory', 192.

O_{n+1} is the offspring of the individual represented by state O_n, or whether O_{n+1} is the same individual as O_n but at a later moment in time. So, whatever the distinction is between biological and cultural evolution, or between vertical and horizontal transmission, I assume a network model to be able to capture all relevant practices that influence our knowledge. What I am interested in, is the evolution of concepts: theories and beliefs that define entities and phenomena that we come across in our practices. That means I *am* making a clear distinction between practices and concepts. In the literature on evolutionary epistemology, this distinction is sometimes blurred, and cultural change is taken to be about both change in beliefs as well as in practices.

I end with a reference of two facets of my model: that of agency, and of bidirectional causation between agent and environment. In this section, I explained that the selective pressures of the environment on an agent cannot be described externally from that agent, because the agent influences its environment through its practices. In chapter 4, I explain how in my model, the environment of an organism can be considered in agential terms.

Chapter 3: The later Wittgenstein and his place in model-building

One problem of building model of the evolution of concepts analogous to the evolution of organisms, is that concepts either must be attributed unrealistic properties to have them behave as organisms do, or that these properties are merely metaphorical. The first would be unconvincing, the second is criticized by Joseph Fracchia and R. C. Lewontin.[57] In this chapter, I argue that my model does not merely function as an analogy, because it uses ideas of the later Wittgenstein to connect the evolution concepts to the evolution of organisms and their practices. The organisms and their practices evolve according to evolutionary theory as discussed in section 2.3. The evolution of concepts is an extension of the evolution of organisms and their practices.

In section 3.1, I address the criticism of Fracchia and Lewontin. In section 3.2, I introduce the necessary Wittgensteinian ideas as tools to explain how the evolution of concepts is an extension of the evolution of organisms. I end this section with the idea of a *pre-theoretic notion* that can be conceptualized. In section 3.3, I give an example of *colour* as a pre-theoretic notion and *colour names* as a possible conceptualization. I end section 3.4 with explaining the exploratory function of my model as a conceptualization of the intuitive notion of a *concept*.

3.1 Not just an analogy

In section 2.1, we have seen that a common approach in the EET program was to create an analogy between the evolution of organisms and the evolution of science, at least for *variation*, *selection*, and *transmission/retention*. In section 2.3, I claimed any analogy needs an account of what it means for a concept to be selected. We have seen that simply saying reality selects for the best theories in the same way that the environment selects for the best organisms gets us into trouble: organisms are not just passive adaptations to an unchanging environment, but are an active part of that environment.

In their paper "The Price of Metaphor," Joseph Fracchia and Richard Lewontin echo some points made in chapter 2. Their main criticism is that any historical development is not "explicable only by the "one force" of selection."[58] They think the selectionist paradigm overlooks the importance of individuals and behaviours that are non-adaptive.[59] Actions of individuals do matter – they are no more vehicles for "fit concepts" than organisms are vehicles for fit genes.[60] On this I agree, and I already addressed these issues in chapter 2. Perhaps an analogy could still be made to work. However, though mostly criticizing selectionist analogies, Fracchia and Lewontin also point out problems with analogies that are more general.

Let me take a step back and consider what I am looking for. I want to build an evolutionary model that explains why the development of knowledge took *this* path rather than another. Fracchia and Lewontin agree that any analysis of historical change should explain why *this* path was taken instead of another.[61] However, they point out that analogous explanations in other domains "can always be made to work, but they don't do any useful work."[62] The simple fact we can think of knowledge as being selected, using

---

[57] Joseph Fracchia and Richard Charles Lewontin, 'The Price of Metaphor', *History and Theory* 44, no. 1 (February 2005): 14–29.
[58] Fracchia and Lewontin, 24.
[59] Fracchia and Lewontin, 26–27.
[60] Fracchia and Lewontin, 29.
[61] Fracchia and Lewontin, 23–24.
[62] Fracchia and Lewontin, 15.

all the same terminology as is used in biology, does not mean that this tells us anything new about the development of knowledge, and it certainly does not mean that knowledge *is* driven by selection. More important than whether the analogy fits the empirical data, Fracchia and Lewontin claim, is whether the assumptions behind the analogy are appropriate to the domain it is used on.[63] Even if knowledge does develop in an evolutionary way, why would it develop in the exact same way as organisms do?[64]

I am sympathetic to this train of thought. Concepts are not the same entities as organisms, and evolutionary theory is built on the properties of organisms. Why then apply that same theory to conceptual change? It might be true that a lot of elements in biological evolution do not work the same in conceptual evolution. Perhaps variation emerges less randomly, and more as a response to specific problems.[65] Perhaps the survival of concepts has less to do with fitness than the survival of organisms does.[66] If we keep pointing out such discrepancies, one might start to wonder what will be left of the analogy.

Does that mean we should just give up on the whole EET program? Forcing ourselves to look at any domain through paradigmatic glasses means we run into the danger of overlooking important elements or ascribing properties to entities that do not have them. This is true for the selectionist paradigm, which is why I rejected it in chapter 2 in favour of a paradigm informed by the Extended Evolutionary Synthesis. But that does not get us out of the woods. Lewontin is not against the use of analogies, but he is of the opinion that it can never be the analogy itself that brings us new knowledge.[67] Models are supposed to be imperfect simplifications, that much is true. However, even though gas particles might behave sufficiently similar to billiard balls to describe them as such, we will make some obvious mistakes if we ascribe the same properties to gas particles as to billiard balls.[68] Lewontin remarks that a billiard ball could only be a valid model of a gas particle *in spite of* the ways in which gas particles and billiard balls differ, not because of them.[69] "[T]he model should be chosen *before* the metaphor and not by means of it. The metaphor is to be chosen by virtue of its elements of similarity to the pre-existent structure of rules [of the model]."[70]

That does not mean that a model cannot have any metaphorical elements, but for Lewontin, the roles that are left for those elements to play are didactic (relating the model to more familiar experience helps us understand it better), experimental (simplified models are easier to apply), or unifying (metaphors as "articles of faith in the unity of experience").[71] Hope for unification has undoubtedly been a driving force behind the EET program. Darwin's theory of evolution seemed so elegantly simple that it was surely worth it to apply it to other domains. Of course, evolution is *not* that simple. But it is a theory *of change*, and we are interested in how and why knowledge changed, so why would we not try to apply it to that domain?

---

[63] Fracchia and Lewontin, 17.
[64] Fracchia and Lewontin, 18.
[65] Fracchia and Lewontin, 21.
[66] Fracchia and Lewontin, 22.
[67] Richard Charles Lewontin, 'Models, Mathematics and Metaphors', *Synthese* 15, no. 2 (1963): 229.
[68] Lewontin, 230.
[69] Lewontin, 228.
[70] Lewontin, 229.
[71] Lewontin, 228–29.

I think Lewontin's worries about overlooking important elements or ascribing properties to entities that do not have them are justifiable, but not unique to metaphors. Whatever theory or model we apply to the development of knowledge, it will inevitably ascribe more importance to some data than to others. "[T]he the best material model for a cat is another, or preferably the same cat," Arturo Rosenbleuth and Norbert Wiener quip.[72] Yet if simply looking at the cat itself could tell us everything we wanted to know about the cat, there would be no need for a model. The things we study using models are too complex to be understood without models. In studying change in knowledge, we must deal with a lot of missing or multi-interpretable historical data.

What we need is to build a model that does not just infer the same properties for concepts as for organisms, solely on the basis of some analogous elements between the two domains. As I put it at the end of section 2.3: we need a way to relate evolution to conceptual change that is different from applying evolutionary theory to conceptual change as an analogy. We need a way to relate concepts to the actions of the organisms. My model does that, by taking the evolution of concepts to result from the evolution of actual people and their practices. To be able to explain what that means, I first need to introduce a philosophical toolbox taken from the later Wittgenstein's ideas on meaning.

3.2 Tools taken from the later Wittgenstein

As this is not a thesis on Wittgenstein, this section is not meant to be a defence of Wittgenstein, or one particular interpretation of him. That is why I stay close to the interpretations found in the *Blackwell companion* to Wittgenstein.[73] I use the later Wittgenstein – as outlined in the *Blackwell companion* – as a philosophical toolbox.

A lot of the relevant Wittgensteinian ideas can be found in Wittgenstein's *Philosophical Investigations*.[74] Central is the idea that "the meaning of a word is its use in the language."[75] The way Wittgenstein thinks language functions is through language-games. Like in a game, correct moves (uses of words) are determined by the *rules* (of the language-game).[76] At the start of *Philosophical Investigations*, Wittgenstein provides a simple example of a language:

> The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose, they use a language consisting of the words "block", "pillar", "slab", "beam". A calls them out; – B brings the stone which he has learned to bring at such-and-such a call.[77]

[72] Arturo Rosenbleuth and Norbert Wiener, 'The Role of Models in Science', *Philosophy of Science* 12, no. 4 (October 1945): 320, https://doi.org/10.1086/286874; quoted in Lewontin, 'Models, Mathematics and Metaphors', 227.
[73] Hans-Johann Glock and John Hyman, eds., *A Companion to Wittgenstein* (Chichester: John Wiley & Sons, 2017).
[74] Ludwig Wittgenstein, *Philosophical Investigations*, ed. G. E. M. Anscombe, R. Rhees, and G. H. Von Wright, 3rd edition (Oxford: Basil Blackwell, 1968).
[75] Wittgenstein, para. 43; quoted in Gary Ebbs, 'Rules and Rule-Following', in *A Companion to Wittgenstein*, ed. Hans-Johann Glock and John Hyman (Chichester: John Wiley & Sons, 2017), 392.
[76] Daniel Whiting, 'Language, Language-Games, and Forms of Life', in *A Companion to Wittgenstein*, ed. Hans-Johann Glock and John Hyman (Chichester: John Wiley & Sons, 2017), 422.
[77] Wittgenstein, *Philosophical Investigations*, para. 2; quoted in Ebbs, 'Rules and Rule-Following', 391.

The rules of this game are to bring a slab when the word "slab" is called out, to bring a pillar when the word "pillar" is called out, etc. How does A teach B these rules? According to Wittgenstein, this teaching "is not explanation, but training."[78] The goal of this training is for A to *understand* "slab" to mean that he should bring a slab to B. We can easily tell if A has properly understood the meaning of "slab", simply by looking whether he actually brings a slab. However, there lies a puzzlement in how we should explain that *that* course of action (the bringing of a slab) is the correct application of *this* rule. Perhaps the builder is following quite a different rule, and is only (correctly) bringing slabs by accident. We feel we need an explanation of how a somehow predetermined rule *guides* us to the correct application.[79] But any explanation we could give, would again be open to further, possibly *incorrect* interpretation. This is Wittgenstein's rule-following dilemma: "No course of action could be determined by a rule, because any course of action can be made out to accord with the rule."[80]

Wittgenstein's 'solution' is to say that "there is a way of grasping the rule which is not an interpretation, but which is expressed in what we call 'obeying the rule' and 'going against it' in actual cases."[81] The only way to talk about correctness is in the context of "rule-following practices."[82] The idea that there are some bedrock facts that predetermine correct applications of rules is mistaken – there are only "our actual practices of *taking* some new applications to be correct and others incorrect."[83]

Remember that we are looking for a way to connect evolution to conceptual change. In the paragraphs above, we can see how meaning can be coupled with *correct application*. Ordinarily, one might say that the meaning of the concept "slab" is that it refers to an actual slab (and not a pillar, block, beam, etc.). But for Wittgenstein, the concept exists because of the language-game between the builder and assistant. Without the need to build with slabs and pillars, why would we want to have these concepts? In this particular language-game, understanding the meaning of "slab" means understanding what to do when somebody calls "slab!". This implies the existence of a rule for the correct application of this concept "slab": e.g., one that tells you to bring a slab if and only if somebody calls "slab!". But Wittgenstein's rule-following dilemma implies such a rule could never be written down, and could never be interpreted. For Wittgenstein, the meaning of a concept can only be expressed in *action*. The meaning of "slab" is expressed in the act of the assistant bringing a slab when being told "slab!".

Games have purposes as well as rules.[84] The purpose of the building game is to build something. In that context, "slab" means something, but in a different context, it might mean something else. That is why, given a certain context, we have concepts for some things but not for others. That is *not* to say that the context *determines* the meaning, but that the context determines whether a language-game is being played.[85] For example, the proposition "I am here" has meaning in the context of a conversation via Skype, where upon saying the words, I show a map and point towards a location. But if I randomly exclaim "I am here," I cannot mean anything by it, if the words are not uttered in a "suitable" situation.[86]

---

[78] Wittgenstein, *Philosophical Investigations*, para. 5; quoted in Ebbs, 'Rules and Rule-Following', 391.
[79] Ebbs, 'Rules and Rule-Following', 398.
[80] Wittgenstein, *Philosophical Investigations*, para. 201; quoted in Ebbs, 'Rules and Rule-Following', 399.
[81] Wittgenstein, *Philosophical Investigations*, para. 201; quoted in Ebbs, 'Rules and Rule-Following', 396.
[82] Ebbs, 'Rules and Rule-Following', 400.
[83] Ebbs, 400.
[84] Whiting, 'Language, Language-Games, and Forms of Life', 424.
[85] Whiting, 427.
[86] Whiting, 427.

Does that mean language has no meaning outside of the intentional context of games? As Wittgenstein puts it, "the speaking of language is part of an activity, or of a *form of life*."[87] There is at least one form of life shared by all humans: our biological constitution – the fact we are all part of the same species.[88] This shared form of life illustrates what for Wittgenstein is the bedrock of our expressions. Not bedrock in any justificatory sense: the rule-following dilemma shows that that is impossible. In what sense, then? When I utter "I am here" without any clear intentional context, Wittgenstein agrees on the *certainty* – the problem is not one of radical skepticism. Wittgenstein just does not think it counts as knowledge precisely because it is not uttered in a language-game.[89] Knowledge, seen as justified true belief, needs some form of grounding, some demonstration of its truth.[90] Such grounding and demonstration can only happen in the context of a language-game, so there is no way to express propositions outside such games.

If knowledge and certainty are not the same, what exactly is the certainty that "I am here", and how is our knowledge based on it? Wittgenstein calls these *basic certainties* "certainties *in action*," functioning as "unjustifiable *rules of grammar*".[91] The fact that "I am here" can be uttered does not give it propositional content. Wittgenstein thinks we cannot be mistaken about basic certainties.[92] The same words might be uttered in a different intentional context (in a language-game) which does give it propositional content, like in the Skype example where one points at a map and says, "I am here." But then any doubt we could have about the truth of that statement would rest on the rules of that language-game.

To be clear, propositions have their place. The proposition "there is a slab," expressed in a room with slabs, is a justified true belief. The question for Wittgenstein is, why is it justified? Because the person saying "there is a slab" only says that in a room where there *really are* slabs? How would we know this to be the case? The rule-following dilemma shows that a person's grasp of the meaning of "slab" cannot be expressed by a rule like "only say 'there is a slab' in a room with *actual* slabs," because it just shifts the problem: how do we know that person knows what *actual* slabs are? Justification presupposes rules, and the rule-following dilemma shows that this process of justification must end somewhere, for if I said, "these foundations are *necessarily* true," I would rightly be asked "but what predetermined rule is there to decide this and how do I get to know it?" That is what basic certainties are for Wittgenstein: the inevitable bedrock of our convictions. Basic certainties are not true or false, but they are the bedrock of all our convictions, true or false.

This does not imply that basic certainties are merely based on induction, as Wittgenstein makes clear: "'The certainty that the fire will burn me is based on induction.' Does that mean that I argue to myself: 'Fire has always burned me, so it will happen now too?' Or is the previous experience the cause of my

---

[87] Wittgenstein, *Philosophical Investigations*, para. 23; quoted in Whiting, 'Language, Language-Games, and Forms of Life', 424, my emphasis.
[88] Whiting, 'Language, Language-Games, and Forms of Life', 424.
[89] Danièle Moyal-Sharrock, 'Wittgenstein on Knowledge and Certainty', in *A Companion to Wittgenstein*, ed. Hans-Johann Glock and John Hyman (Chichester: John Wiley & Sons, 2017), 547.
[90] Moyal-Sharrock, 548.
[91] Moyal-Sharrock, 549.
[92] Moyal-Sharrock, 549–50.

certainty, not its ground?"[93] Elsewhere, he gives the answer that is it is indeed the cause, but not the ground: "The squirrel does not infer by induction that it is going to need stores next winter as well. And no more do we need a law of induction to justify our actions or our predictions."[94]

The main takeaway is that expressions have meanings only within the intentional context of language-games. The only way to talk about correct use of an expression is within such context, by reference to the rules of the *language-game*. The *rule-following* dilemma shows us that we can attempt to determine why *this* application of the rule is the correct one – in other words, that we can attempt to ground the meanings of the expressions used in the game – but that we will never reach justificatory bedrock. At bedrock, there is only our *use* of expressions, our actual practices and activities in which we employ them. This bedrock might be expressed in the form of *basic certainties*, but these certainties themselves cannot be justified or unjustified in the same way an expression in the intentional context of a language-game can be justified.

So, according to this theory of meaning, our practices make us play certain language-games, which results in the existence of certain concepts. We are organisms that build things, and therefore have a purpose for the simple building language outlined in the beginning of this section. That language-game includes rules for interpreting the different concepts like "slab" and "pillar," but should one ask "why are we justified in believing that the concept 'slab' really refers to slabs," we could only go so far as to demonstrate this meaning by the rules of the language-game. The rule-following dilemma shows that for the question "why *these* rules" we cannot provide a ground – we can only point at the practices that made us play this game in the first place. This is what I meant when at the end of section 3.1 I said that the evolution of concepts results from the evolution of actual people and their practices.

How should this reliance of concepts on practices be understood? In *Philosophical Investigations*, Wittgenstein rejects the Platonic view that every concept has an essence.[95] Even a concept with seemingly clear definitions like 'number' has no essence. We call something a number because it resembles other things that have previously been called number.[96] All forms of numbers – rational numbers, real numbers, complex numbers – form a *family*, connected by a web of *resemblances*. The boundaries of the concept 'number' *could* be defined mathematically, but that is not how we *use* the concept. In fact, our use of the concept 'number' was extended when we invented (or discovered) complex numbers.[97] Rules for correct use of 'number' (for the right *definition* of 'number'), then, come *after* the use, not before. It is *family resemblance* that connects particular uses of the same word.

Different practices result in the emergence and change of different family resemblance concepts. Just so is our family resemblance concept of 'number' connected to our mathematical practices. In other words, practices change the language-games in which we use concepts. Then, we *can* (but do not have to) make up rules that support us in our application of these concepts, in the form of definitions. Any entity or

---

[93] Wittgenstein, *Philosophical Investigations*, para. 325; quoted in Moyal-Sharrock, 'Wittgenstein on Knowledge and Certainty', 553.
[94] Ludwig Wittgenstein, *On Certainty*, ed. Denis Paul and G. E. M. Anscombe (Oxford: Basil Blackwell, 1969), para. 87; quoted in Moyal-Sharrock, 'Wittgenstein on Knowledge and Certainty', 554.
[95] Hanoch Ben-Yami, 'Vagueness and Family Resemblance', in *A Companion to Wittgenstein*, ed. Hans-Johann Glock and John Hyman (Chichester: John Wiley & Sons, 2017), 407–19.
[96] Ben-Yami, 411.
[97] Ben-Yami, 414.

phenomenon X can be defined by a rule: by describing the exact circumstances in which one should identify the entity or phenomenon as X. We can do this either implicitly, or explicitly in the case of formal theory, which allows us to apply a concept more purposefully. This way, this application informs our practices, which can then result in new changes to our family resemblance concepts.

To summarize: family resemblance concepts form the connection between practices and theoretical concepts. They should be seen as *pre-theoretic notions* that are invoked by practices, and forcefully present themselves to us, even if we cannot define them. Anyone that has had experience with e.g., cats or numbers, has a family resemblance notion of 'cat' and 'number' in mind. Having studied mathematics – having engaged in certain mathematical practices – my concept of 'number' might be broader than that of somebody who has not done so, because I have been involved in more practices in which things played a role that resemble what I have previously called a number. Collectively, we use these concepts in language-games. We can, and in science generally do, make up rules to define these concepts. These are then replicated and *evolve* through use in language-games.

I now introduce a second facet of my model: the connection between the bidirectionally connected agents I introduced in chapter 2, and the concepts that, in the Wittgensteinian way explained in this section, result from the practices of those organisms. These practices have purposes, possibly unknown to the organism, and the practices can be expressed in the form of basic certainties, which merely reflect the practices but have no propositional content themselves. The practices can be modelled by considering agents and their activities and interactions with each other and their surroundings. These activities and interactions result in the emergence (and change) of certain family resemblance concepts in the organisms that have concepts: us. We may define rules for the correct application of these concepts. This gives rise to theory, which can then be applied. Through this application, these concepts are tested, and through this application they influence the way agents interact with each other and their environment – in other words, influences our practices.

3.3 The pre-theoretic notion and conceptualization of *colour*

In this section, I discuss an example of a pre-theoretic notion and one conceptualization of that notion: colour, and its categorization into discrete terms.

What is colour? What we are seeing is the spectrum of visible light, different colours corresponding to different electromagnetic waves. This visible spectrum is a consequence of evolution. Given these constraints, we employ different practices that make use of our ability to see colours. We cannot make use of ultraviolet lights like birds do, but we are able to recognize red berries in a green bush. In fact, we think about the whole spectrum in discrete terms and call many different electromagnetic waves 'red,' for example. Colour is a pre-theoretic notion. It is a consequence of the practices in which we make use of colours. Our categorization of the visible spectrum into names like 'red' and 'blue' is a way to formalize this notion. For most people, 'red' is connected to 'pink' and 'orange,' but perhaps more closely connected to 'maroon,' which people are likely to call 'red' in some circumstances. How does such a categorization arise, and is it the same for everyone?

Part of the answer is of course that humans have certain photoreceptor cells in our eyes that respond differently to different (length, uniformity, and height of) electromagnetic waves. But from this biological fact seemingly does not follow one unique discrete categorization of the colour spectrum. Nevertheless, empirical data suggests that there is some universality in the (development of) colour categorization in different languages. There is a theory by Brent Berlin and Paul Kay that posits that (1) there exists a set of primary colours – black, white, red, yellow, green, and blue – that form the basis of the colour categorization of most languages in the world, and (2) languages evolve this colour categorization in a somewhat fixed order – starting with black and white, then red, then yellow and green, and then blue.[98] The World Color Survey provides some empirical evidence for Berlin and Kay's model in the form of colour-naming data from speakers of 110 unwritten languages.[99] Gibson et al. analysed this data and found that warm colours (reds and yellows) are communicated more efficiently than cool colours (greens and blues).[100] Their explanation for Berlin and Kay's second hypothesis is that "[t]he use of color terms depends on communicative needs," and that because "[o]bjects (what we talk about) are typically warm-colored, and backgrounds are cool-colored," languages evolve terms for warm colours before they evolve terms for cool colours.[101] Berlin and Kay's own explanation – in line with the hypothesis that use depends on needs – is that in "technologically simple" societies, colour is a less informative descriptor than in "technologically complex" ones, because as skill in dyeing increases, some objects might only be distinguishable by colour – e.g. the same piece of clothing being sold in multiple colours.[102]

In both explanations, we can see the influence that practices have on our pre-theoretic notion of colour, and therefore on our subsequent categorization of it. The theory is that we have concepts for warm colours before we have concepts for cool colours, because the need to distinguish and talk about warm colours arises first – as a consequence of, say, our practice of picking red berries on the green backdrop of the forest. Similarly, our practice of dyeing has supposedly given rise to an even more elaborate partitioning of the visible spectrum of light. We could always see the same spectrum, but what the pre-theoretic notion 'colour' *means* has changed because our practices have changed.

Note that by definition, a pre-theoretic notion cannot be put into words. It may therefore seem confusing to talk about what the pre-theoretic notion 'colour' means, because meaning is something that we think can be written down. My claim is that we all have an idea of what 'colour' is before having written anything down, even though this idea is now hard to separate from the more formal conceptualizations of 'colour' we have in mind – our categorization into 'red,' 'blue,' 'green,' etc. being one of those conceptualizations. A formalization of a pre-theoretic notion into a concept is just a way to talk about the notion, and the point is that practices have influenced the notion of 'colour' enough to give rise to new ways of talking about the notion.

3.4 The pre-theoretic notion of *concept*

In this thesis, I assume that concepts, and conceptual change, are suitable to represent all (scientific) thought and theory. I explained their link with practices, and made the distinction between pre-theoretic notions and theoretical concepts. Concepts vary within the context of the practices that gave

---

[98] Paul Kay and Luisa Maffi, 'Color Appearance and the Emergence and Evolution of Basic Color Lexicons', *American Anthropologist* 101, no. 4 (1999): 744, https://doi.org/10.1525/aa.1999.101.4.743.
[99] Paul Kay and Richard S. Cook, 'World Color Survey', in *Encyclopedia of Color Science and Technology*, ed. Ronnier Luo (Berlin, Heidelberg: Springer Berlin Heidelberg, 2015), 1–8, https://doi.org/10.1007/978-3-642-27851-8_113-10.
[100] Edward Gibson et al., 'Color Naming across Languages Reflects Color Use', *Proceedings of the National Academy of Sciences* 114, no. 40 (3 October 2017): 10788, https://doi.org/10.1073/pnas.1619666114, fig. 4.
[101] Gibson et al., 10785.
[102] Kay and Maffi, 'Color Appearance and the Emergence and Evolution of Basic Color Lexicons', 746.

rise to their pre-theoretic notions, and within a network of other notions that are connected to it through meaning. Going from pre-theoretic notions to theoretical concepts, however, is not straightforward. There are degrees of formalization, a continuum of conceptual clarity. It seems difficult to give a formal concept of 'concept.'

However, if we take Wittgenstein and the model seriously, we should not be concerned about this. 'Concept' itself is a notion, and this thesis is just one formalization of that notion – specifically in the domain of (history of) logic. This application is an exploration of an alternative to the Platonic view, rejected by Wittgenstein, that notions have 'true' essences which we are slowly but surely discovering.

Whatever practices the pre-theoretic notion of 'concept' is involved in, it is perfectly coherent for there to be multiple formalizations of this notion: one of which the Platonic view, the other the Wittgensteinian view. Whether my Wittgensteinian formalization will *survive* – I am using this word deliberately, because I am talking about conceptual *evolution* – depends on whether it makes the practices it belongs to more successful. The two formalizations are only competing insofar they reject each other, but there is no reason they cannot survive simultaneously, at least temporarily, just like there is no reason different practices in the same environment cannot survive simultaneously if both these practices are replicated in other organisms somehow.

The advantage of my Wittgensteinian formalization, as opposed to the Platonic view, is that it allows room for the peculiarity that there is not necessarily a right answer to the question of what a concept is. Just like one organism could convince the other that their practice is the superior one, so too am I convincing the reader of the superiority of my view of the concept "concept" – except what we think of as "convincing the other" is really an evolutionary process, and the survival of one organism, or practice, or concept, does not make it 'better,' but at best 'fitter' *in a certain context*. For practices, this context is the natural world in which organisms carry out the practices. For concepts, it is the pre-theoretic world of basic certainties which the practices have given rise to. The Wittgensteinian view is self-coherent in this way.

Of course, a Platonist might object to all of this, and instead say that the essence of 'concept' is for concepts to have an essence. Perhaps that view is also self-coherent. However, the important difference between the Platonic view and my view is that my view can explain why the Platonic view exists, while the Platonic view can only say about my view that it is wrong. My thesis is just one replication of the Wittgensteinian, evolutionary formalization of the notion 'concept,' and it might die out because it fails to make our practices more successful. How this works, I explain in more detail in section 4.4.

This also shows one important purpose of my model: to provide a conceptual framework to talk about the notion of 'concept,' and more generally, about knowledge. We can also find this purpose of modelling in ecology. Ecology often makes use of mathematical models, with mathematical equations that are supposed to represent ecological processes and concepts. Jay Odenbaugh points out that these concepts have lives that are independent of their mathematical counterparts.[103] As an example, he refers to Robert May, who explored the connection between complexity and stability.[104] May wondered what happens to the equilibrium of an ecological community when it is perturbed, and whether more

[103] Jay Odenbaugh, 'Idealized, Inaccurate but Successful: A Pragmatic Approach to Evaluating Models in Theoretical Ecology', *Biology & Philosophy* 20, no. 2–3 (March 2005): 245–46, https://doi.org/10.1007/s10539-004-0478-6.
[104] Odenbaugh, 246.

complex communities were more stable. May's modelling revealed multiple ways to think about complexity and stability.[105] This resulted in multiple definitions for complexity and stability, leading to a large number of contending ways to think about the hypothesis that more complex communities are more stable.[106]

May's modelling has helped make sense of the phenomenon of complexity and stability. We have some pre-theoretic notions of complexity and stability even without being able to define them clearly, because we have experience with dynamic processes and networks. In other words, our practices that involve these dynamic processes have given rise to the pre-theoretic notions of stability and complexity. Even if the model never stated definitively that this is how complexity and stability *really* work, it still provided us with a new framework to talk about these issues. So too does my model provide a new framework to talk about conceptual change, about the pre-theoretic notion of 'concept.'

Taking my own model seriously as a conceptualization of the pre-theoretic notion of 'concept,' means that it should serve the purpose of getting grip on a practice – some philosophical or scientific practice of theorizing. According to my model, all knowledge is, is different attempts at conceptualizing practices. The problem of the methods that focus on drawing analogies, as I explained in section 2.1 and criticized in section 3.1, is that these methods do not take place *in* a practice, and therefore have no meaning. At best, they identify the fact that multiple practices can be linked by featuring the same or similar pre-theoretic notions: e.g., of variation, replication, and selection.

In the next chapter, I develop my conceptualization of the pre-theoretic notion of *concept*, and, by extension – following Wittgenstein as discussed in section 3.2, and the EES as discussed in section 2.3 – the pre-theoretic notions of *practice* and *organism*. The model should be seen as an extension of evolutionary theory, rather than an analogy of it.

[105] Odenbaugh, 249.
[106] Odenbaugh, 250.

Chapter 4: The model

In this chapter, I develop my model of the evolution of concepts as an extension of the evolution of practices and organisms. My model can best be seen as the conceptualization of three pre-theoretic notions: of *organism*, *agent*, and *concept*. I explain three distinctions which follow from these notions, by considering what is external to each notion and what is internal to it. This forces us to distinguish the inorganic from the organic (section 4.1), passive attributes from active practices (section 4.2), and demonstrations from concepts (section 4.3). In section 4.3, I show a visualization of the model. In section 4.4, I end this chapter with a discussion of how the notion of *correctness* should be considered from the perspective of my model.

It is not my intention to claim of anything that it is really on one side of a distinction or of the other. I merely identify three pre-theoretic notions, and claim that given the framework of evolution, these three distinctions follow from these pre-theoretic notions. It is also not my intention to claim that one side of a distinction is in any way prior to the other. For example, to say that there is a distinction between the organic and the inorganic, is *not* necessarily to say that the organic can be explained in terms of the inorganic.

The model is best explained in the order in which it has been built up, starting with the three distinctions. This order of explanation is purely conceptual, and not indicative of the historicity of any of the notions addressed. However, since practices are performed *by agents*, and concepts are grounded in practices, one would first need to assume that organisms exist before being able to talk of practices, and – given Wittgenstein's theory of meaning as explained in section 3.2 – one would first need to assume that practices exist before being able to talk of concepts.

4.1 Distinguishing the organic from the inorganic

As I mention in the introduction of this chapter, my intention is not to make hard claims about what an organism is, how it is related to the inorganic, or whether and how inorganic material can become organic. The way I talk of organisms in this section, is in an intuitive sense: I *am* claiming there is some pre-theoretic notion of *organism*, from which follows a distinction between the organic and the inorganic.

How might we describe this intuitive sense of the organic? When looking at the world, we notice an assortment of molecules and processes. Some of these molecules and processes seem to form a stable system that actively preserves and advance its own existence. We call these systems 'organisms' and distinguish them from *inorganic* materials and processes. We might not be able to precisely define what it means for molecules to be organic (leading, for example, to discussions about whether we should call viruses organisms), but we generally recognize organisms when we see them. These organic systems are made up out of inorganic material, implying some form of gradation between the inorganic and the organic – although this does not mean that the inorganic is necessarily prior to the organic in any way.

We can make the distinction between the inorganic and the organic without needing reference to evolutionary theory. I deliberately say "*we can make* the distinction", because it is not in principle *impossible* to describe all matter purely in terms of chemical or physical processes between molecules or atoms. However, to repeat, I do not intend to claim that the inorganic is necessarily prior to the organic: that the organic in the end *really is* inorganic, or that it is explainable in terms of the inorganic. I merely claim that we distinguish the organic from the inorganic because it explains the world better than not doing so.

This is the position I started from in chapter 2. Evolutionary theory sets out to explain change in lineages (populations) of organisms, and the adaptationist stance does so by claiming there is variation in each organism, which is heritable, and which exhibits differential fitness. In this view, organisms can be seen as sets of attributes that have developed randomly, constrained by the environment so as not to hurt the organism's self-preservation and continued existence.

Very crudely, we can consider a lion as having attributes like strong muscles, a good sense of smell, a response to low energy intake that includes chasing antelope, and a desire to procreate with other lions. It has attributes *like these* because by definition, it actively preserves its own existence, and therefore needs energy intake. It has *these* attributes because it got them from other organisms in its lineage, because they mutated randomly, and because the environment that it lives in, consisting of inorganic matter and other organic systems, allowed for these attributes to be passed on to the lion by other organisms in its lineage with similar sets of attributes.

Having distinguished organic, self-preserving systems from the rest of the world, we must demarcate that which is part of the *internal* processes of the organism, and that which is *external* to it. The two are interdependent: the internal processes exist in and interact with the external. Energy intake – the process of eating – serves the *internal* purpose of satisfying the organism's appetite, but that appetite developed because it aided in, or at least did not hurt the survival of the organism's predecessors in the *external* environment.

In this view, which forms the basis of my model, organisms are passive vehicles in the process of evolution.[107] All that they do are responses to stimuli. Some of these responses survive and are propagated to other organisms, others die out. The responses can be complicated, so complicated that we might not always fully understand why and how they work, but however complicated, they are still attributes of the organism that developed because of random variation and independent selective pressures.

4.2 Distinguishing active practices from passive attributes

In section 2.3, I discussed the Extended Evolutionary Synthesis (EES), which gives organisms an *active* causal role to play in evolution. In this view, a lion chasing antelope is not just a response to a hunger stimulus, but an active choice the lion makes to deal with its hunger. Not all that the lion is and does, is *active*.

This idea of active behaviour rather than passive stimulus responses comes from a pre-theoretic notion of *agency*. I do not intend to make any specific claims about agency, except that granting agency to organisms means making the distinction between the *passive/reactive* parts of the organism, and the *active* parts. Again, I deliberately say "*making* the distinction" because it is not in principle *impossible* to consider all parts of an organism as passive. Like we distinguish the organic from the organic because it explains the world better than not doing so, the EES distinguishes the active from the passive parts of an

[107] It is not my intention to claim that this is the view of proponents of the adaptationist stance. In this section, I merely use my simplified conception of the adaptationist stance as a starting point to explain organisms as carriers of passive attributes.

organism because it claims that this explains the world better than not doing so.[108] Both distinctions are *chosen* for their increased explanatory power.

Without a strict definition of *organism*, we might not always be able to say what (parts of) organisms we should consider to be organic, and what (parts of) organisms we should consider to be inorganic. Similarly, without a strict definition of *agency*, we might not always be able to say what (parts of) organisms are active, and what (parts of) organisms are passive. Intuitively, I might say of myself that I actively choose the words I speak, but that I am not actively breathing. For the practice/attribute distinction, as for the organic/inorganic distinction, there exists a gradation.

I make the specific distinction between what I call active *practices* of the organism, and its passive *attributes* – between what it has control over *as an agent*, and what it does not have control over. For example, an organism seemingly has no control over the fact that it experiences hunger (experiencing hunger is an *attribute*), but it does seem to have control over how it responds to this experience (chasing antelope is a *practice*). An organism seemingly has no control over the fact that it has legs (having legs is an *attribute*), but it does seem to have control over how it uses these legs (walking and running are *practices*). To repeat, it is not in principle *impossible* to construe a practice as an attribute: we can consider chasing antelope to be an active choice of the lion, or a passive response to stimuli. Whether these examples of a lion's attributes and practices are really correct is beyond the scope of my thesis. All I require for the case study in chapter 6, is the assumption that the logicians I discuss there are indeed organisms that partake in some active practices of reasoning.

What kind of active practices can organisms have? Like the properties of organic systems are *constrained* by the inorganic environment, the practices are constrained by the attributes of the organism: like an organism that lives in an underground inorganic environment cannot develop wings, an organism that has only legs (as attributes) cannot partake in the practice of flying. Furthermore, like the properties of organic systems are constrained by the processes of *other* organic systems, the practices of an organism are constrained by the practices of *other* organisms: the antelope-chasing practice of the lion is adapted to the lion-evading practice of the antelope (and vice versa).

In this view, informed by the EES, organisms can be envisioned as interdependent *agents*, causally connected through their practices. The lion and the antelope are connected because of the antelope-chasing practices of the lion and the lion-evading practices of the antelope. Organisms can use their attributes for multiple practices. Organisms with similar sets of attributes can use these in different practices: one lion might chase antelopes by isolating the weakest, another might use the benefit of surprise to its advantage. We generally group both activities into one practice with some internal variation, like we group organisms with similar attributes into one population or species with some internal variation. Practices form lineages just as organisms form lineages: a practice might be copied by one organism from another organism and then varied slightly, like offspring will have a similar but slightly varied set of attributes compared to its ancestors.

When we consider the evolution of organisms, we generally do not scrutinize individual organisms, but groups or lineages of organisms that have similar attributes. These groups exist by virtue of having proven that their set of attributes *works* in the environment they live in: one individual set of attributes

[108] See section 2.3.

would not have been replicated multiple times if it was not suited to the environment. So, when we speak of fitness of organisms, or of their survival, or of their replication, we are interested in which group – which set of attributes – is better *given the context of a certain environment* (consisting of the inorganic and the other groups of organisms). The idea of evolution being context-dependent is key. There is no way to speak of fitness, or survival, or replication of organisms, if they do not make use of and are not constrained by the same environmental context somehow. Are the attributes of a whale better than those of a mayfly? In terms of the amount of time it allows them to live, perhaps, but for the mayfly that has no concept of time longer than a day or two, why would that be relevant? But the whale *is* better at living in the ocean, which poses dangers, constraints, and opportunities very different from the environment of the mayfly. There is a reason that the whale has fins, and the mayfly has wings.

The context of a group of organisms also has to do with the other groups that they are causally connected to. Antelopes cannot develop and propagate any attributes or practices if they are killed by lions. In section 2.3, I discussed Baedke et al.'s conception of this connectedness. In figure 1 – an example of niche construction of coral – the two connected entities were chosen to be the coral and its environment, but such a network can be as elaborate as one wants. As Baedke et al. put it, "[t]he precise interpretation of the states and relations depends on the causal narrative a scientist is putting forward, given a particular research case."[109] The choice for specific states and relations are always an abstraction from the totality of connected organisms and inorganic environments. The key takeaway is that the evolutionary context for coral is different than for a human, and that therefore, humans have developed different attributes, practices, and concepts.

We can see now how we might talk about practices in a similar way as we talk about organisms. We should investigate the fitness, survival, and replication of practices in the same way. We consider groups of self-similar practices that have proven they work: one practice would not have been replicated multiple times if it did not work. In the case of practices, what is the context in which they do or do not work? There is no way to speak of fitness, or survival, or replication of practices if they do not make use of and are not constrained by the same set of attributes. Can we say that the practice of flying is better than the practice of walking? In terms of speed, perhaps, but for an organism that does not have the attributes to fly at all, why would that be relevant? Such an organism can only develop practices that are possible within the constraints of their set of attributes.

When we consider the evolution of organisms, we *fix the context of a certain environment*. We do so by making the distinction I made earlier, between what is *internal* to the organism and what is *external* to it, and by considering how the *internal* changes in the fixed *external*: how do internal sets of attributes evolve given certain external conditions? Of course, the external conditions themselves also change, both the inorganic and the other organic systems. For example, the amount of oxygen and radiation in the atmosphere has changed over time, which has eventually allowed organisms to move from the water and exist on land. But, again, sets of attributes can only be compared and evaluated given a fixed context. When we talk about the fitness, survival, and replication of one group of organisms, we construe the other organisms as part of the same inorganic background. That is what it means to fix the context.

---

[109] Baedke, Fábregas-Tejeda, and Prieto, 'Unknotting Reciprocal Causation between Organism and Environment', 14.

When we consider the evolution of practices, we again *fix the context*. This time, the context is the set of attributes that constrained and made possible the practices under consideration. Again, we make the distinction between what is *internal* to the agent and what is *external* to it: how do its practices evolve given a certain set of attributes, and given the existence of other, related practices that influence it? Everything external to the practice can of course also change both the attributes of other organisms as well as the other related practices. But when we talk about the fitness, survival, and replication of the practices of one agent, we construe the practices of the other agents as simply being their attributes, essentially removing their agency.

This might seem like a strange step, but first remember that we *can* in principle view practices as mere attributes. To make sense of the necessity of this step, consider how we would explain the best practices – the best moves – of players in a game of chess. Each player has certain attributes: the pieces they have, and how they can move them in later turns. Their practices are the choices they make in response to a certain state of the game, the moves they make given a certain turn. Such choices do not only depend on the pieces on the board, but also what the one player thinks the other player will do in response. In that sense, the practices of both players are important. However, we can only consider the *agency* of one player at the same time. When we consider the practices of player A, we reduce the practices of player B to the choices player A thinks they will make. Player A might think that if she moves her knight, player B will move his pawn with 20% chance, and rook with 80% chance. The agency of player B has been removed from this perspective. His practices have been reduced to a mere attribute: as a response to the stimulus of seeing player A's knight moved, he will move his pawn with 20% chance, and his rook with 80% chance. Player B does not really move his pieces randomly like that, but from the perspective of the best practices of player A, he might as well.

One might object that chess moves cannot happen simultaneously, while organisms do not have to wait for each other to act. Suppose we consider the practice of the lion chasing antelopes. Even if the antelope is capable of changing its evasion practices while being chased, when we consider what chasing practices of the lion are best, we can only do so *given a fixed context*, given a fixed way that the antelope reacts. The antelope might have different evasion practices, and to know what chasing practices are best, we need to take into account all of them. What I mean by fixing the context is not to say that the rest of the organisms in the network do not react at all, or could not react in a multitude of ways. It means that whatever ways they *can* react in, does not change. The practices of the other organisms do not disappear, but become attributes, mere facts of the organisms that make them react a certain way to stimuli like a machine, rather than possible *actions* that an agent can perform and adapt.

To summarize, we consider the evolution of organisms – construed as sets of attributes – in a certain fixed inorganic context. When we study one group of organisms, we fix its context by abstracting away the evolution of the other groups of organisms with which this group is causally connected. In other words, other organic systems have been made part of the world external to the organism under consideration, merely seen as yet more unchanging inorganic selective pressures. That is what Baedke et al. do when they construe all the species that coral interacts with as part of the same environmental entity.[110]

[110] See section 2.3, figure 1.

Similarly, we consider the evolution of practices of organisms – construed as agents – in the context of a fixed set of attributes. When we want to study one type of practice of one group of organisms, we fix its context by abstracting away the practices of other organisms. We do that by construing these other practices as mere passive attributes, which nevertheless influence the active choices of the organism that is being studied, thereby influencing the selective pressures on its practice.

These two *evolutions*, of organisms and of practices, are shown in figure 3 together with the evolution of concepts, which I discuss in the next section. Each evolution can be considered separately from the evolutions to its right, but not separately from the evolutions to its left, because I consider the practices *of* organisms, and because concepts are grounded in these practices, in the Wittgensteinian way explained in section 3.2. These dependencies are visualized in the ovals: if organisms have the necessary level of agency, then we can consider the evolution of their practices, and if agents have the necessary level of intelligence, we can consider the evolution of their concepts.[111]

4.3 Distinguishing demonstrations from concepts

The next step is to consider those agents that have the capability (the intelligence) of knowledge – or concepts, in my models. These concepts get their meaning from the practices in which they are used. Before I further explain this dependency of concepts on practices, I explain how practices are related to each other.

Some practices, like walking and running, are more similar than others, like running and flying. We generally do not consider every individual instance of a practice separately, like we did not consider every individual organism separately. We group organism by virtue of being similar and coming from the same lineage – in other words, by having been replicated from one another. So too do we group practices by virtue of being similar and having been replicated from one another. For example, practices are copied by students from their teachers, by children from their parents, or by peers from their peers. When we speak of the practice of running, we generally like to ignore small differences between the ways individual people run. This is no different than what we do when we speak of a lion: we ignore the small differences between individual lions.

Given such a group of practices, for which the internal variation is marginal, and for which we fix the external context made up of other organisms with their own practices, we can consider the evolution of the concepts of the agents partaking in these practices. We want to be able to compare concepts, which is only possible if we can compare them within the same fixed context, within the same practice. Otherwise, any variation between concepts might be the result of it being grounded in a different practice.

Remember that we could distinguish the passive attributes of an organism from its active practices. Now, I make another distinction, between the parts of a practice that can merely be *demonstrated*, and the parts that can be *conceptualized*. Where the distinction between attributes and practices was the result of the agency with which we let the EES imbue organisms, the distinction between demonstration and concept comes from Wittgensteinean theory of meaning as discussed in chapter 3. Wittgenstein said that the bedrock of our knowledge – what he called *basic certainties* – cannot be true or false,

[111] I use intelligence merely to refer to the capability of having concepts, or knowledge.

because they are mere demonstrations of our practices. Yet, within the context of a practice, we can and do create concepts.

The practice of counting can be demonstrated, e.g., by use of an abacus, but it can also be conceptualized, for example by giving the mathematical definition of a number and of operations like addition, multiplication, etc. Like parts of an organism can in principle be explained either in terms of attributes or in terms of practices, parts of a practice can in principle be explained either in terms of demonstrations or in terms of concepts. Again, there is a gradation between a demonstration and a concept. What I have called a pre-theoretic notion is that familiar experience of a demonstration that seems to have one essence, the way that it seems impossible to count something and not think of the pre-theoretic notion of ‘number’ – regardless of any specific conceptualization of numbers.

These transitional phases also exist in the earlier two distinctions I have made, between the inorganic and the organic, and between the passive and active parts of an organism. We can visualize these three distinctions in the following schema in figure 2.

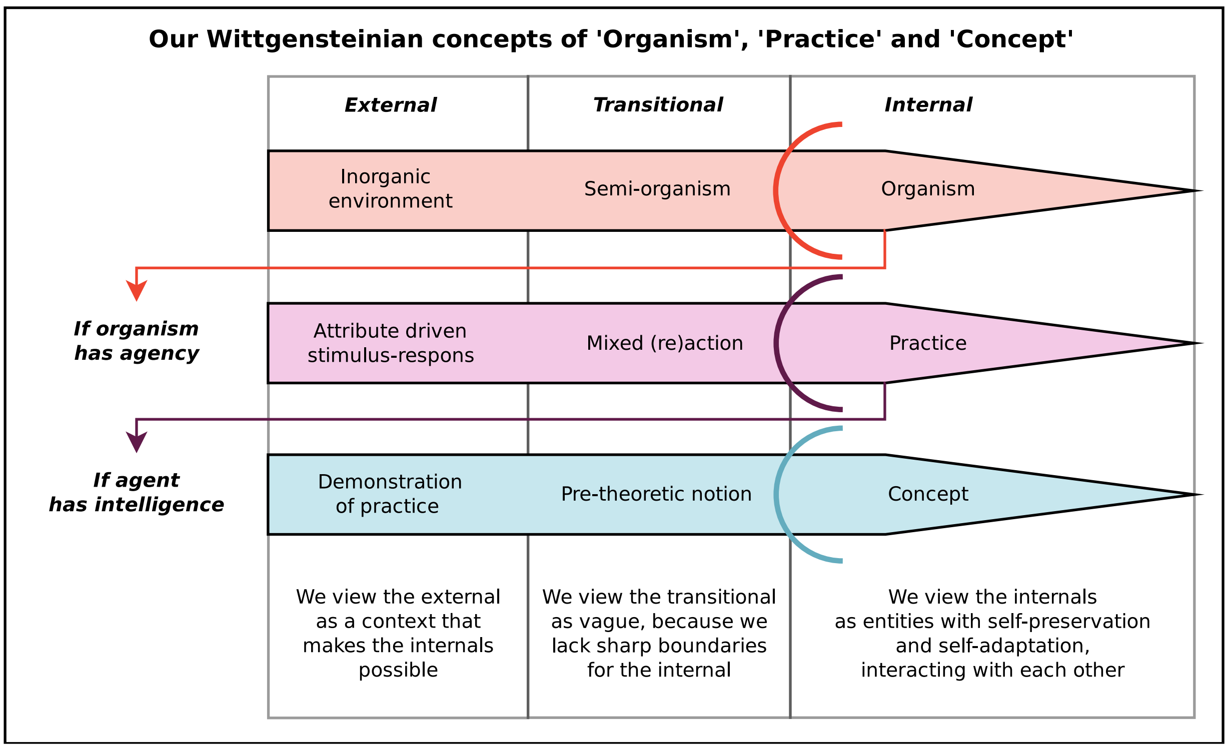

Figure 2. A schematic overview of the three distinctions discussed so far, between the organic and the organic, between passive attributes and active practices, and between demonstrations and concepts. Organism, practice, and concept are called Wittgensteinian concepts because the model itself is a conceptualization of these three pre-theoretic notions.

‘Transitional’ here does not mean that one side of the distinction *becomes* the other, but that anything can in principle be explained in terms of either side of the distinction, and that there may exist situations in which this choice is not straightforward.

In section 4.1, I claimed that we *deliberately* distinguish the organic self-preserving systems from the rest of the inorganic world because it explains the world better than not doing so. I also claimed that this forces us to demarcate what is *internal* to the organism, and what is *external* to it. Because we lack sharp boundaries for the internal, there exists a transitional phase between the two, *semi-organisms* like viruses, of which we are not fully sure whether they should be explained in organic or in inorganic terms.

In section 4.2, I claimed that we *deliberately* distinguish the actions that an organism chooses to make *as an agent*, which can be construed as *practices*, and the *re*actions that it exhibits as a response to stimuli, which can be construed as *attributes*. Again, this forces us to demarcate what is internal to the agent and what is external to it. Because we lack sharp definitions for what parts of what organisms have agency, there exists a transitional phase of *mixed (re)actions*, behaviour of which we are not sure whether it is best explained as the result of agency or as passive stimulus responses. Consider the way people often talk of animals, for example. Intuitively, a dog seems to have more agency than an ant, but less than a human. I am not claiming that this is true, but clearly there is some gradation when it comes to agency. Even the allegedly prototype agent, the human, has little to no active control over its beating heart.

In the current section, I consider concepts, which for Wittgenstein, as I have discussed in chapter 3, are connected to practices. I claim that we deliberately distinguish those parts of practices that can merely be demonstrated, which Wittgenstein called *basic certainties*, from those parts that can be conceptualized. Again, this forces us to demarcate what is internal to the concept and what is external to it. Because we lack sharp boundaries for what parts of our practices can be conceptualized, there exists a transitional phase of *pre-theoretic notions* which are more than mere demonstrations and are capable of being conceptualized, though imperfectly.

I claimed that the practice of counting can be both demonstrated and conceptualized. Any demonstration or any conceptualization invokes the experience of the pre-theoretic notion of *number*. By definition, a pre-theoretic notion cannot be put into words, because the choice of words would already conceptualize it in some way. We merely use the word ‘number’ to refer to the experience of that notion which the practice of counting invokes. Pre-theoretic notions form the basis of our conceptualizations, but a concept is not a one-on-one translation of a pre-theoretic notion – it is not meant to replace it. It is at once an improvement of a pre-theoretic notion – by making it sharper and more applicable – as well as it is an impoverishment of one, because it is not a direct translation.

Having made these distinctions between external and internal for all three domains, we can visualize the evolutions of organisms, practices, and concepts in the schema below in figure 3.

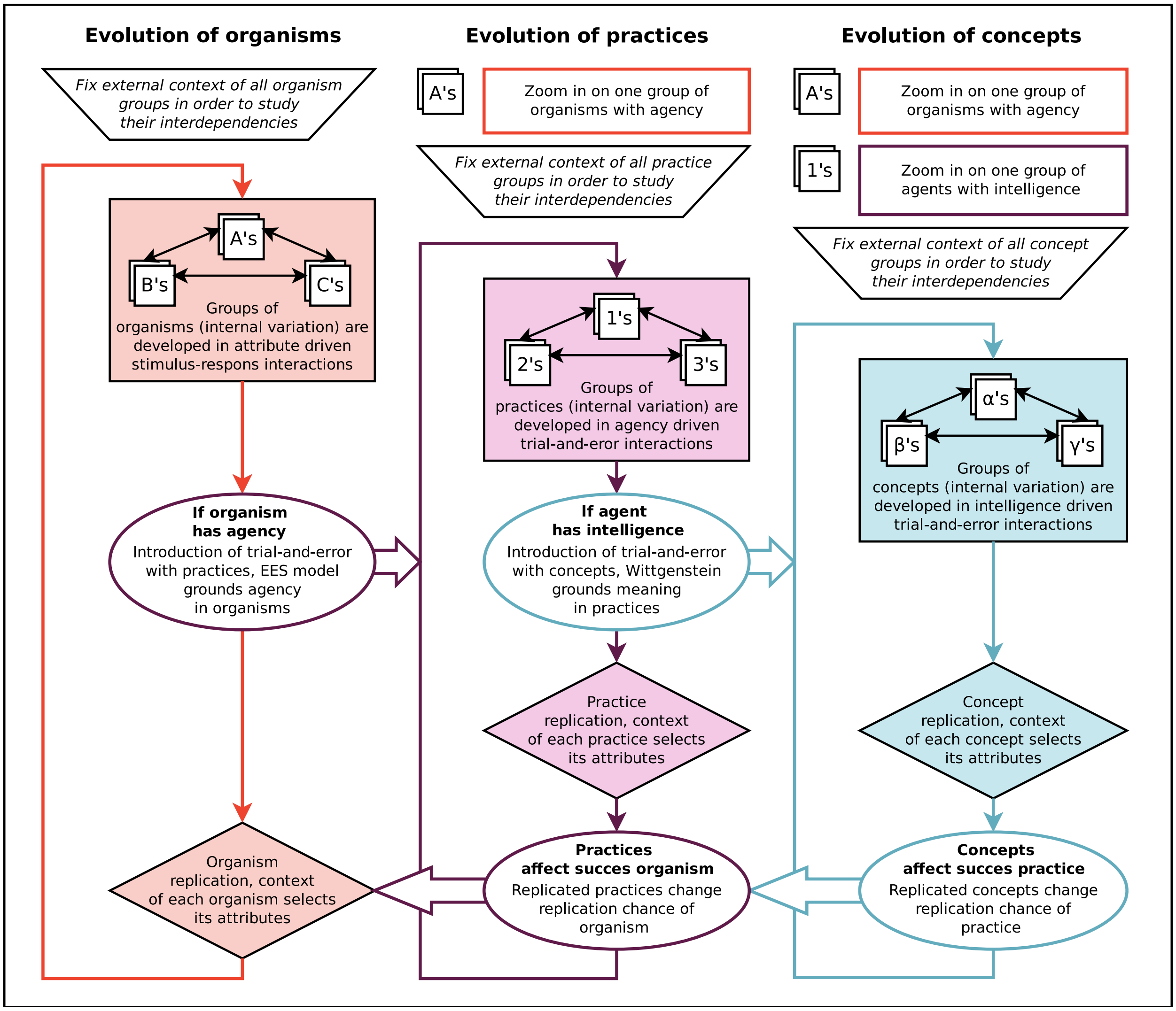

Figure 3. A visualization of the model, consisting of three evolutions. The colours of each evolution correspond to the colours used in the distinction schema of figure 2.

The schema in figure 3 not only shows the internal variation-selection-retention loop of each evolution separately (as indicated by the small arrows), but also two external loops, *between* evolutions (as indicated by the bigger arrows). Consider as an example the development of *clap skates* – ice skates that have the boots attached to the blade at the toe, only by one hinge, allowing the blade to maintain contact with the ice for a longer time. We can see the clap skate as a conceptualization constrained by the practice of, say, gliding over ice. This practice invokes a pre-theoretic notion of *something to glide with* – be it a sleigh, a pair of skis, or a pair of ice skates. Even if we narrow down the practice to ice skating in particular, there are various ice skate designs that make us perform the practice differently. These different designs are different conceptualizations.

So, one external loop starts with a practice (of gliding or skating) that constrains the possible designs or conceptualizations with which this practice can be performed: it has to be something that can glide effectively. Given the intention behind the practice (e.g. moving around), some designs will glide more effectively than others, and will for that reason be more likely to be used by other agents. The success of the practice itself – whether it is replicated – will also depend on the conceptualizations used: a skate that helps the agent get around faster will make the agent more likely to choose the practice of skating rather than some other practice of moving around, like walking or cycling. The practice of skating gave rise to the clap skate, which in turn *changed* our way of skating.

The other external loop starts with a set of attributes that constrains the possible practices that can be done with it. Like the design of the clap skate is constrained by the practice of skating, so too is the practice of skating constrained by the attributes of the agents partaking in it: we evolved to have these legs with all their muscles and joints, and given those legs, we evolved different ways to use them in practices, the practice of skating being one of them. But a practice can also improve us as organisms, by making it more likely for us to replicate. For example, the practice of skating can make it easier for us to get around in winter, which improves our chances of survival in environments where winter gets cold enough for water to freeze.

4.4 The issue of correctness

Before moving on to the case study in logic, I need to make a note about the issue of correctness. Given any pre-theoretic notion, one might think there is a measure of correctness that says which concept best describes this pre-theoretic notion. But, as I show in this section, this rests on a misunderstanding that can be solved by adopting the Wittgensteinian perspective from section 3.2.

Pre-theoretic notions originate in experiences with practices. The pre-theoretic notion of a 'hammer' presupposes the practice of hammering – or, to put it more generally, the practice of inserting a slender object into some substance by hitting it with a blunt object. Driving a nail into wood with a hammer is hammering, but so is driving a tent peg into the ground with a rock. Even though we now call it *hammer*ing, the practice of hammering came before the object 'hammer.' Any specific type of hammer is just one formalization of the pre-theoretic notion of a hammer. If we never had the practice of hammering, we would not be talking of hammers, because we would have no experience with this pre-theoretic notion.

The notion of a hammer can only be made comprehensible by using a hammer. Why even bother coming up with formalizations, then? Because they *improve* the practices. Formalizations of hammers – in this case, perhaps better called designs – improve the practice of hammering not by some objective standard of proper hammering, but by creating the opportunity for the practice, carried out with that specific hammer design, to more easily be replicated.

One might wonder why for each practice, we do not have the perfect concepts – in other words, for each practice, why has not just one concept replicated itself: the best, most correct one? How would we speak of the *correctness* of a concept? Defining one concept into another will not do, because, as Wittgenstein claimed, we will never reach justificatory bedrock.[112]

As an analogy, consider how we would speak of the best, most correct practice, given a certain goal. We come up with practices because they improve us *as organisms*. A more correct practice is one that results in more improvement. Improvement should not be understood as some objective measure, but as related to intentions. The lion chases the antelope because it experiences hunger. The practice of

[112] See section 3.2.

antelope-chasing improves the lion because it satiates the lion, and because it did not hurt its chances of survival, it can freely be replicated.

But why should this not be the only practice that is replicated, if it serves the goal of satiating the lion so well? Because practices serve multiple goals, not all of which we are always aware of. Consider our own different practices of eating. At any given meal, we can choose what we eat, but we cannot eat everything. Different choices will not only satiate hunger to different extents, but also serve other purposes like happiness and healthiness. For example, we have to choose between the practice of eating fast food, and the practice of eating vegetables. Eating fast food satiates hunger and makes us happy in the short term. Eating vegetables also satiates hunger, but to a lesser extent. It does make us healthier and happier in the long term. If we merely consider the goal of satiating hunger, we can compare both practices by referring to their ability to satiate hunger, and might declare fast food the winner. Nevertheless, because both practices serve multiple different goals, they can both be replicated in our society.

The same analysis applies to concepts. When we use a word, it derives its meaning from the practice we use it in. But words can derive their meaning from multiple practices. That we use words differently in different practices is unproblematic in itself, as long as we can be sure of what practice we are using it in at any given moment. When I refer to 'truth' in a philosophical practice I mean something different than when I say something is true in an informal conversation. As long as the other is carrying out the same practice, there is no issue. The problem is that when we are not sure what practices we are carrying out, we can unknowingly refer to 'truth' in the context of one practice at one moment, and in the next moment refer to it in the context of another practice. And the more abstract practices become, the harder it is to keep track of which practice we are in exactly – which goal we are trying to fulfil. Even within the philosophical community, different concepts of truth get replicated and co-exist because there are different overlapping practices at play.

In the next two chapters, the challenge will be to show this is the case for logical concepts like logical consequence.

Chapter 5: The legitimacy of logical theories from the perspective of the model

This chapter serves the purpose of introducing logic – both as a (reasoning) practice and as a conceptualization of this practice. In section 5.1, I start by explaining what logic is about. Any such explanation will involve different intuitive senses of the notion of logical consequence: put more strongly, logic is *about logical consequence*. I discuss six notions of logical consequence in section 5.2, taken from Stewart Shapiro's paper "Logical Consequence, Proof Theory, and Model Theory."[113]

This raises questions about the legitimacy of logical theories: which of these notions are they supposed to capture, and when can they be said to be correct, or at least more correct than some other theory of the same notion? The answers to these questions are beyond the scope of my thesis, but they are a good starting point to consider how my model handles matters of legitimacy differently than logicians usually do. In section 5.3 and 5.4, I discuss the ideas of Stewart Shapiro and Graham Priest about the matter of legitimacy of logical theories. For this, I use their papers "Logical Consequence, Proof Theory, and Model Theory," and "Revising Logic," respectively.[114] It is not my intention to disprove Shapiro or Priest. Rather, I use their thoughts on the matter as a way to explain potential misconceptions of my model. It also serves the purpose of fine-tuning the model for an application to logic in particular.

5.1 What is logic about?

What is logic about? According to Stewart Shapiro and Kouri Kissel, logic intends to capture *correct reasoning*.[115] This has generally been the subject-matter of philosophy. Nowadays, logic has been thoroughly formalized, and is often subsumed under mathematics. This raises the question: how do the properties of this formalization relate to what logic, at least originally, was about? If we consider logic purely as a mathematical curiosity, this question does not have an answer, because mathematics need not be *about* anything. But historically and philosophically, there is something pre-theoretic which formal logic is a theory of. The main pre-theoretic notion I am interested in, is the notion of *logical consequence*: in an argument, in what sense do conclusions deductively follow from their premises?[116]

Shapiro identifies six notions of logical consequence that he claims can be attributed to various logicians in history.[117] All of these have shortcomings: the choice for one of them is not unambiguous. The two modern concepts of logical consequence, the proof-theoretic and the model-theoretic definition, are based on some of these six notions. A discussion of these six notions will show that there is such a thing

---

[113] Stewart Shapiro, 'Logical Consequence, Proof Theory, and Model Theory', in *Oxford Handbook of Philosophy of Math and Logic* (Cary, UNITED STATES: Oxford University Press, Incorporated, 2005), 651–70, http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3052016.
[114] Graham Priest, 'Revising Logic', in *The Metaphysics of Logic*, ed. Penelope Rush (Cambridge: Cambridge University Press, 2014), 211–23, https://doi.org/10.1017/CBO9781139626279.016.
[115] Stewart Shapiro and Teresa Kouri Kissel, 'Classical Logic', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2021 (Metaphysics Research Lab, Stanford University, 2021), sec. 1, https://plato.stanford.edu/archives/spr2021/entries/logic-classical/.
[116] Jc Beall, Greg Restall, and Gil Sagi, 'Logical Consequence', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2019 (Metaphysics Research Lab, Stanford University, 2019), https://plato.stanford.edu/archives/spr2019/entries/logical-consequence/.
[117] Shapiro, 'Logical Consequence, Proof Theory, and Model Theory'.

as a pre-theoretic notion of logical consequence – or perhaps multiple. In chapter 6, the goal is to show that our choice of these concepts depends on the practices that we use them in.

All six notions are about *deductive* consequence, meaning that there is some idea of *necessity* involved in how the conclusion follows from the premises. This is unlike *inductive* consequence, where the conclusion follows not necessarily, but *likely* from the premises.[118] It is not in principle wrong to argue that all swans are white because you have only seen white swans up until this point, but it is not a deductive argument, and therefore not a logical one. There is still the possibility that the next swan you will see, is black. In this thesis, I concern myself with logic and logical consequence, and therefore with deductive arguments.

The conclusion and premises of a deductive argument consists of sentences. Like Shapiro, I denote these sentences with capital Greek letters (e.g. Φ and Γ). In a formal logical system, the distinction between formal sentences (that the system uses in arguments) and natural sentences (that we use in arguments) becomes important.[119] Insofar such a system is supposed to represent correct reasoning, the formal sentences represent their natural counterparts. However, it is not in principle impossible for there to be properties of formal sentences that natural sentences do not have, or vice versa. As I said at the beginning of this section, a formal system does not necessarily have to be *about* anything. But if logic hopes to be normative, then it must be about something. This pre-theoretic something, specifically the pre-theoretic notion of logical consequence, is what I discuss in this chapter, and we can hence take the capital Greek letters to refer to natural sentences.

Another preliminary note is that by definition, it is difficult to talk about pre-theoretic notions without formalising them in the process. Writing them down – using words to express them – can be seen as a formalisation. All six notions could be, and sometimes have been, formalizations of logical consequence. However, the way I discuss them here is supposed to show that they intend to capture something pre-theoretic.

5.2 Six pre-theoretic notions of logical consequence

The first two notions Shapiro introduces, have to do with the necessity I pointed out earlier:

1. "Φ is a logical consequence of Γ if it is not possible for the members of Γ to be true and Φ to be false."[120]
2. "Φ is a logical consequence of Γ if Φ is true in every possible world in which Γ is true."[121]

In these two notions, Φ refers to the conclusion, and Γ to the premises.[122] In both notions, necessity is expressed in terms of impossibility or modality. The notions are mostly similar, but the terminology of possible worlds is the more modern choice. Clearly, these notions describe some form of consequence, but do they also describe *logical* consequence?

Consider the following sentences:

---

[118] Beall, Restall, and Sagi, 'Logical Consequence', sec. 1.
[119] Shapiro, 'Logical Consequence, Proof Theory, and Model Theory', 652.
[120] Shapiro, 655.
[121] Shapiro, 655.
[122] An argument can and often does have multiple premises, but in most logics it can only have one conclusion. A capital Greek letter can refer to either a single sentence, or to a set which includes one *or more* sentences.

Γ. B is taller than A

Φ. A is shorter than B

Does the conclusion Φ follow from the premise Γ? According to these first two notions of logical consequence, it does: there are no possible worlds in which Γ is true but Φ is false. This is because of the meanings of 'taller' and 'shorter.' If Φ is false, then A is not shorter than B. But if Γ is true, then B is taller than A, so A cannot fail to be shorter than B. That is a contradiction.

However, the general view is that logical consequence holds by virtue of *form*.[123] The argument above holds because of its *content*: what 'taller' and 'shorter' mean *matters* for the truth of the argument. One would be right to point out that there are more words than just 'taller' and 'shorter' whose meaning play a role in the validity of the argument. Though debates about this topic exist, the prevailing view is simply that some of these words are *logical*, and others are not. Specifically, the connectives "not", "and", "or", and "if… then", the quantifiers "some" and "all, and some variables (e.g. capital letters) are called *logical* terms.[124] An argument is true only if it is true by virtue of the meanings of only *these* terms, not others. Put in another way, it means that the meaning of any nonlogical terminology used in the argument can change, and the argument would still hold true.

The importance of meaning is captured in the next two notions Shapiro introduces:

3. "Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the terms in those sentences."[125]
4. "Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the logical terminology."[126]

Notion (3) captures meaning, but not specifically *logical* meaning, and therefore does not solve the problem of the example involving length. As I have said, we say that an argument holds by virtue of form (as opposed to content) if the only meaning that is important is the meaning of the *logical* terminology. This is captured by notion (4). Note that this is not a necessary requirement for a logic to have.

Another idea important to logical consequence is that of normativity. If logic is about *correct* reasoning, the rules of a true logic should be followed if we want our arguments to be valid. This is captured by the last two notions of logical consequence that Shapiro discusses:

5. "Φ is a logical consequence of Γ if it is irrational to maintain that every member of Γ is true and that Φ is false. The premises Γ alone *justify* the conclusion Φ."[127]
6. "Φ is a logical consequence of Γ if there is a deduction of Φ from Γ by a chain of legitimate, gapfree (self-evident) rules of inference."[128]

Notion (5) assumes rationality of subjects that use reasoning. That is not to say that they *always* act rational, but that, if they reason *irrationality*, we say that they "*could have known better*, and indeed

[123] Shapiro, 'Logical Consequence, Proof Theory, and Model Theory', 657.
[124] Shapiro, 659.
[125] Shapiro, 656.
[126] Shapiro, 659.
[127] Shapiro, 659.
[128] Shapiro, 660.

should have known better," as Shapiro puts it.[129] That normativity is captured by the rules mentioned in notion (6). If the rules mentioned are *correct*, then notion (5) follows from notion (6), because it is irrational to not reason according to correct rules.[130]

5.3 The matter of legitimacy from Shapiro's perspective

According to modern standard logic, there are two formal definitions of logical consequence: a proof-theoretic and a model-theoretic one. The proof-theoretic notion is the formal version of notion (6).[131] It captures the idea of normativity and rationality, but note that it is the only notion that does not use the concept of truth. The model-theoretic notion corresponds to a combination of notion (1), (2), and (4): "Φ is a logical consequence of Γ if Φ is true in every possible world under every reinterpretation of the nonlogical terminology in which every member of Γ is true."[132] It captures the idea of modality, truth, and logical form.

The proof-theoretic defines truth in terms of logical consequence, while the model-theoretic gives conceptual priority to truth by defining logical consequence in terms of truth.[133] In either case, truth and logical consequence are connected. The notion of logical consequence clearly involves a myriad of other notions: normativity, rationality, modality, truth, meaning, form, etc. If we want a correct formalisation of logical consequence – in other words, if we want to definitively define what *correct* reasoning is – we should also come up with correct formalisations of these other related concepts. Even if we assume that this is possible, these concepts themselves would also make use of each other and of other concepts. We would still be missing *criteria* for this correctness: criteria for when the rules mentioned in the proof-theoretic notion are correct, and criteria for the truth mentioned in the model-theoretic notion.

Shapiro admits he cannot give an account of how we are supposed to tell that the pre-theoretic notions behind the formalisations are legitimate.[134] Part of this confusion stems from Shapiro using terms like "notion," "conception," and "definition" interchangeably. For example, he introduces the six notions as being different ways of thinking about the "intuitive, or pretheoretic [*sic*], notion of logical consequence."[135] Later on in the same paper, he calls notion (5) "another *definition* of consequence."[136] Anyone that claims that a logic is a theory *of* something, must admit that there is a distinction between the pre-theoretic notions, and the formalizations of those notions. As I have said in section 5.1, the fact that we must use words makes it difficult to discuss pre-theoretic notions while keeping them distinct from formalisations, but the difference is important.

It seems, then, that if we claim that logic *is not* a theory *of* something, we get rid of the issue of legitimacy – of how to determine correctness. Note that, as I have said in section 5.1, there is no reason we could not study any formal mathematical system *qua (pure) mathematics*. On this view, mathematics

---

[129] Shapiro, 660.
[130] Shapiro, 661.
[131] Shapiro, 661.
[132] Shapiro, 663.
[133] Peter Schroeder-Heister, 'Proof-Theoretic versus Model-Theoretic Consequence', in *The Logica Yearbook 2007*, ed. Michal Peliš (Prague: Filosofia, 2007), 187–200.
[134] Shapiro, 'Logical Consequence, Proof Theory, and Model Theory', 667.
[135] Shapiro, 654.
[136] Shapiro, 660, emphasis mine.

is the study of formal systems.[137] We could study formal logical systems not as systems that model correct reasoning, but as systems in and of themselves, with mathematical properties that relate to each other in different ways. In fact, this is an interesting enterprise even if we *do* think that logic is about correct reasoning. For example, the proof-theoretic and model-theoretic formalisations of logical consequence relate to each other in ways that do not *necessarily* have anything to do with correct reasoning.[138]

Would that get rid of the demand for correctness? No, the problem is more persistent than that. The way I construe practices, pure mathematics, even in the view that the formal systems it studies are not *about* anything, is a practice too. 'Formal system' is not an empty term, but finds meaning in its use by mathematicians. Even if we take the view that all formal systems are equal, the study of formal systems by mathematicians unavoidably identifies some things as formal systems, and other things as not. In fact, this view is contested precisely because, in practice, mathematicians do *not* treat all formal systems as equal.[139]

I agree with Shapiro that these two formal notions intend to capture different pre-theoretic notions of logical consequence, and in that sense, there is no issue of legitimacy: they can and do co-exist.[140] Suppose model theory and proof theory are about two different practices. Within the context of one practice, then, one would expect there to be one 'best' concept of logical consequence. However, as I explained in section 4.4, when we use a label like 'logical consequence,' we are not always sure in the context of which practices we are using it. Questions of legitimacy of a concept stem from confusion about the practices that the concept derives its meaning from. That is why Shapiro is unable to explain the legitimacy of the proof-theoretic and model-theoretic concepts of logical consequence.

5.4 The matter of legitimacy from Priest's perspective

Graham Priest points out that questions of legitimacy and correctness of concepts also raise questions of revision: proponents of the canonical logic – classical logic – often claim that logic cannot be revised, and that therefore classical logic is the correct logic.[141] Even from Shapiro's perspective as I adopted it in the beginning of this chapter, this seems a strong claim. From that perspective, whatever logic is about perhaps cannot change, but our theory of it can. By studying logic, we are discovering the laws of correct reasoning in the same way as we discover natural laws by studying physics. The laws are fixed, but our theory is not.

[137] Leon Horsten, 'Philosophy of Mathematics', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2022 (Metaphysics Research Lab, Stanford University, 2022), sec. 2.3, https://plato.stanford.edu/archives/spr2022/entries/philosophy-mathematics/.

[138] See for a short summary Shapiro, 'Logical Consequence, Proof Theory, and Model Theory', 661–69 For any logic, if correctly deducible arguments are also true, the logic is said to be sound. In other words, if the proof-theoretic notion of consequence holds for an argument, so should the model-theoretic notion. Vice versa, a logic is called complete if true arguments are also correctly deducible. Modern standard logic is both sound and complete. Matters that are debated, and that do relate to correct reasoning, are whether soundness is more important than completeness, and whether a logic failing to be sound or complete is really a failure of the arguments it holds correctly deducible, or of the arguments it holds true.

[139] Horsten, 'Philosophy of Mathematics', sec. 2.3.

[140] Shapiro, 'Logical Consequence, Proof Theory, and Model Theory', 667.

[141] Priest, 'Revising Logic', 211.

I now outline Priest's views on the matter, who not only distinguishes between logic as a theory, and what it is a theory *of*, but introduces a more Wittgensteinian notion: *how* we reason logically in practice. These three senses of 'logic' Priest calls *logica docens*, *logica utens*, and *logica ens*.[142] Logica docens is the logic that is taught, logica utens is how people reason, and logica ens is what actually is valid reasoning. Priest's inquiry is whether any or all of these three senses can be revised, and whether they can be revised *rationally*.[143] Comparing Priest's senses to Shapiro's analysis, what do these three senses correspond to? At first glance, logica docens corresponds to the theory and concepts we use in logic. Priest claims that logica ens corresponds to what logic is a theory *of*.[144]

Then what is logica utens? It is not a descriptive matter of psychology, as Priest points out.[145] Rather, it is a normative notion. Priest illustrates this using an example of a reasoning test called the Wason Card Test.[146] The test uses of a pack of cards, each card with a letter on one side and a positive number on the other. The cards are laid out on the table with either the letter or the number face up. A subject is told that "if there is an A on one side of the card, there is an even number on the other."[147] What cards should the subject turn over *at minimum* to verify this statement? A common answer is to turn over the cards that show the letter A and the cards that show an even number. But this is incorrect. The fact that all cards with the letter A also have an even number on the other side, does not mean that all cards with an even number also have the letter A. Checking cards with the letter A might falsify the statement (if they showed an uneven number), but it could never verify it, as there might still be cards with an uneven number that, when turned over, showed the letter A. This is a necessary, but not a sufficient step. Checking cards with *even* numbers is not necessary. It could show cards with a different letter than A without violating the statement, because the statement does not claim that even numbers are exclusive to cards with the letter A. The only way to verify the statement is by turning over all the cards with the letter A *and* all the cards with *uneven* numbers. If a card with an uneven number showed the letter A, then the statement would also be falsified.

When explained, subjects recognise their mistake. Clearly, Priest claims, there is some normative power at play here, something that makes one answer wrong and the other correct.[148] One possible source of this normativity is of course our theory of logic, logica docens. The explanation teaches the subject the correct theory of logic, and therefore results in the recognition that a mistake was made. However, it turns out that even subjects with experience in logical theory get it wrong.[149] Subjects do not recognize their mistake after explanation because of a belief in the correctness of the theory behind the explanation, but because the correct method has been *demonstrated* to them. In my view, that is what logica utens encapsulates: the normativity of the practice of reasoning.

This example displays the often confusing difference between theory and demonstration. It illustrates the Wittgensteinian idea that meaning comes from practices, rather than theory. Priest recognizes this link, and brings up the later Wittgenstein as I discussed him in section 3.2. He wonders, taking

[142] Priest, 212.
[143] Priest, 211.
[144] Priest, 216.
[145] Priest, 219.
[146] Priest, 218.
[147] Priest, 218.
[148] Priest, 219.
[149] Priest, 218.

Wittgenstein seriously, how can our logica utens change?[150] His answer: by training, or repeated practice. However, he moves away from Wittgenstein again when he goes on to claim that clearly, some practices are better than others, and that we choose them rationally based on our best logica docens.[151]

So, when it comes to rational revision, for Priest logica docens is prior. Can logica docens be revised rationally? The answer is not uncontested, but seemingly straightforward. Logics have what Priest calls a "canonical application": deductive reasoning.[152] Theorizing about logic requires "articulation of other important notions, such as truth, meaning, probability."[153] Like with any theory, revision of logica docens is possible "by the standard criteria of rational theory choice."[154]

What are these rational criteria? Priest mentions adequacy to data as the most important one: "those particular inferences that strike us as correct or incorrect."[155] As is the case with any subject matter, both theory and data can be wrong, so more criteria like simplicity are necessary. But here, Priest mixes up data with demonstrations of practices. Taken as data, the inferences subjects make in the Wason Card Test can be wrong for many reasons. But their inferences are always perfect demonstrations of their own practice, even the choice to turn over the cards with even numbers – it is just instrumental to some other goal. Given the goal of verifying the statement outlined at the beginning of the test, subjects *can* be demonstrated the correct reasoning.

As proof that logica docens can be revised rationally, Priest gives an example of one of Aristotle's syllogisms, which I will discuss in more detail in the coming chapter. The syllogism – a kind of logical argument – goes like this: all As are Bs; all As are Cs; therefore, some Bs are Cs.[156] According to Aristotle, the conclusion "some Bs are Cs" follows logically from the premises "all As are Bs" and "all As are Cs." According to modern logic, it does not. We can conclude, then, Priest says, that our logical theory has been revised rationally by modern logicians to make at least this particular inference invalid.[157]

From Priest's perspective, this syllogism should be considered a measurement of data of correct inferences, and according to modern logicians, Aristotle made a mistake in this measurement. The syllogism works if we assume there is at least one A. According to the first premise, that A must be a B. According to the second premise, that A must also be a C. Then that A is both B and C, so the conclusion is correct: there is at least one B that is a C. However, if there are no As, the syllogism fails, as without As, the premises tell us nothing about the existence of Bs and Cs. For whatever reason, Aristotle did not take this possibility into account, which is wrong according to modern logic.

If we do not consider the syllogism as a measurement but as a demonstration of a reasoning practice, then the differences between Aristotle and modern logicians become clear. Consider the following experiment involving a kid, 10 pieces of paper, 3 black stickers, 10 blue ones, and 10 red ones. The kid is

[150] Priest, 219.
[151] Priest, 220.
[152] Priest, 215.
[153] Priest, 216.
[154] Priest, 216.
[155] Priest, 217.
[156] Priest, 213.
[157] Priest, 214.

told that they can divide the stickers over the pieces of paper, but that they must abide by the following rules:

1. All stickers should be used
2. At most *one* black sticker per piece of paper
3. Next to each black sticker should also be a blue *and* a red sticker

When the kid is done, they are asked to pick out the pieces of paper with red *and* blue stickers. Because there are 3 black stickers, at least 3 pieces of paper will have both blue *and* red stickers. The black, blue, and red stickers correspond to the categories A, B, and C in the syllogism, which is validated by a demonstration of the above practice. Within this practice, there is no doubting the demonstration.

From the perspective of modern logic, the existence of black stickers (of As) is not a given, and therefore, the syllogism is invalid. The kid would not be confused if you were to repeat the demonstration, but tell them that now, the black stickers cannot be used. Then there would be no guarantee that there are any pieces of paper with both blue and red stickers (that there are some Bs that are Cs), as the kid might group all blue ones and all red ones together. But what would the kid think if, during the first demonstration, with the first set of rules, there would be no black stickers at all? They would wonder why they were told rule (2) in the first place.

Clearly, the possibility of empty sets is crucial in this example, and I discuss this in more detail in the next chapter. Is it more rational to assume the possibility of empty sets, or is it not? The rationality behind both Aristotle's and the modern logicians' theory is easier understood once you imagine the practice behind it - the practice of actually *using* empty sets, as demonstrated by the sticker example. Considering the whole syllogism to be invalid just because it does not work when there are no As – when there are no black stickers – was irrational for Aristotle, because who would even want to use the syllogism, if there were no As? Who would want to use rule (2), if there were no black stickers?[158] The point here is that demonstrations cannot be wrong in and of themselves. Modern logicians did make a rational choice to make this syllogism invalid, as Priest claims, but so did Aristotle to make it *valid*, and for the same reasons: to improve the respective reasoning practice.

As I have said, Priest claims that our logica utens (our practices) are chosen rationally based on our best logica docens (our best theory). However, as the example above illustrates, what Priest takes for data that the logica docens is based on, I claim is better understood as a demonstration of a practice. If the process is circular, as Priest admits, where is the required normativity supposed to come from?[159] The issue is that Priest misinterprets Wittgenstein. The training of a practice on the basis of a given theory is not the same as the training of a theory on the basis of a given practice. In the latter case, we are not forced by the theory to a certain practice, but the success of the practice forces us to a theory.[160] Recognition of mistakes in the Wason Card Test is not based on the authority of the current logica

[158] In chapter 6, I will claim that it has to do with the genesis of empty sets in a practice: whether the set has always been empty, or has become empty. For the kid, rule 2 makes sense even given an empty set of black stickers, as long as they had experienced the practice while the set was *not* empty first. Rule 2 is only confusing if the set of black stickers had been empty from the start.
[159] Priest, 'Revising Logic', 223.
[160] Success *in evolution* – see chapter 4.

docens, but on the success of the demonstration. Knowledge of logica docens plays no role. Logica docens has not created the practice, it is based upon it.

Does the interpretation of correct inferences as data solve the issue? The data that our logica docens is supposed to be based on, according to Priest, should find its ground in what he called the *logica ens* – what our logica docens is a theory *of*.[161] Priest claims that whether logica ens can change depends on what logical consequence is, exactly – is it something that can change, or not?[162] But then, we are left with the problem of having to find out what the correct way to think about logical consequence is. That is an impossible enterprise without a powerful enough theory, which would involve concepts that find their ground in logica ens or some other *ens*, of which we could ask the same questions all over again.

From the modern mathematical perspective, logical consequence is about the relationship between abstract objects. Priest claims that in this view, the truth about logical consequence – logica ens – cannot change, because the truth of claims about mathematical objects does not change.[163] He admits, however, that even though the truth of claims about logical consequence might not change, we express these claims in language – and if the meanings of words can change, expressions *can* change their truth values.[164] Meaning is involved in logical consequence, argues Priest – namely the meaning of logical connectives – and if we change our theory, this changes at least our interpretation of these meanings.[165] But can it also change the actual meanings? Based on his own interpretation of Wittgenstein, Priest thinks so: "if one revises one's theory, and then brings one's practice into line with it, in the way which we noted may happen, then the usage of the relevant words *is* liable to change. So, then, will their meanings – assuming that meaning supervenes on use (and some version of this view must surely be right)."[166] Here, Priest admits that practices can *change* meanings, without considering the possibility that practices ultimately *give* meaning. If the former is possible, could not whatever process is behind this change, be an explanation for the latter? My model is a proposal for such an explanation.

Priest's reasoning is still circular: our logica docens is chosen depending on the logica ens, our logica utens gets its rationality from our logica docens, and the logica ens can change via our logica utens.[167] This, he rightfully claims, is not problematic in itself. We can only revise the knowledge that we already have. The problem is that Priest's reasoning lacks grounding outside of the rational choice for the best logica docens. What are the criteria for the best logica docens? How and where do we find those criteria? It is the purpose of the next chapter to show that the change in logic follows the change in specific reasoning practices, that in their demonstrations provide a grounding for the pre-theoretic notion of logical consequence and our subsequent formalization. Only when we have a practice that demonstrates correct reasoning, do we want a theory of correct reasoning.

161 Priest, 'Revising Logic', 216–17.
162 Priest, 221. Priest talks about validity rather than logical consequence, but I assume that these two refer to the same notion.
163 Priest, 221.
164 Priest, 221–22.
165 Priest, 222.
166 Priest, 222.
167 Priest, 223.

Chapter 6: Applying my model to logic

Investigating the evolution of logical concepts raises the question of how to identify them in historical works. Naturally, I can only make use of whatever has been transmitted via written works. I have to assume that whatever the logicians I discuss, wrote about logical concepts – or whatever other people wrote about those logicians – is representative of what a community of scholars thought about these concepts at the time. What has been written down will concern concepts, that range from hardly more than a pre-theoretic notion to formal definitions, all of which presuppose a practice that can demonstrate their meaning.

The main issue will be to prevent being anachronistic. As we are investigating the *evolution* of logical concepts, it makes sense for those concepts to be vague at times. But we cannot let our own knowledge of and ideas about logic influence what we think people in history knew and thought about it. The concepts I discuss are general enough to make plausible the idea that these already existed in some form in the time of Socrates. Particular details and distinctions – e.g. between proof-theoretic and model-theoretic consequence – will have to be addressed if they become relevant in the evolution of that concept, but not before.

An agent, or group of agents, do not necessarily have to be aware of the practice that demonstrates the meaning of their concepts. If the purpose and success of a practice is clear, it is not likely to be questioned and can easily be overlooked. So, for example, the practice that Aristotle bases his logic on, cannot simply be reconstructed from what he has explicitly said on the matter. That is why, in each of the following sections, I identify the practice in advance – by way of hypothesis – and then provide arguments that make plausible that it was this practice that led the respective logician to his logical theory. I do not make use of all of the respective logician's work, but focus on the choices that for them, seemed clear and unquestionable.

I have chosen to focus on two key logicians in the history of logic: Aristotle (384 – 322 BC) and Chrysippus (280 – 207 BC). These are the central figures in two crucial moments in the evolution of logic in ancient history, during the beginnings of logical theory. It should be remarked that the used sources are incomplete and from second hand. Historians tend to disagree on correct interpretation. In this thesis, I therefore limit myself to the broad threads supported by authoritative sources.

In the conclusion of the thesis, I address some potential further research on Gottlob Frege (1848 – 1925), an important figure in the mathematical formalization of logic, who I would claim integrates the two practices that Aristotle and Chrysippus built their respective logic on.

6.1 The focus on logical consequence

As I have said, I focus on the notion of logical consequence. In section 5.1, I explained what logic is about. I claimed that historically and philosophically, there is something pre-theoretic which logic is a theory of, and that the main pre-theoretic notion I am interested in, is that of logical consequence: in a valid deductive argument, in what sense do conclusions follow from their premises?

In section 5.3, following an examination of some intuitive notions of logical consequence, I identified a multitude of related notions, like normativity, rationality, modality, truth, meaning, and form. For all logicians, there are multiple notions, that together with logical consequence form one coherent whole in the context of a practice, that demonstrates their meaning. These extra notions have to do with the usability of the practice. In the following sections, I address four of such notions: *proof*, *truth*, the *entities* that can be true or false, and the *meaning* of those entities. A full account of the evolution of logic should analyze all these notions. In this thesis, I limit myself to linking the notion of logical consequence with the other four notions.

6.2 Aristotle

First, I identify and discuss the practice of giving counterexamples to generalizations. Second, I explain how this practice arose and existed before Aristotle. Third, I provide some arguments to make plausible that it was this practice that led Aristotle to his logical theory. Lastly, I provide a suggestion for the link between the notion of logical consequence, and the notions of *proof*, *truth*, *entities* of truth, and *meaning* – as explained in section 6.1 – and which I already preview in the next section.

6.2.1 The practice of giving counterexamples to generalizations

The idea behind this practice is simple: if person X claims something is always the case, or is never the case, and person Y demonstrates the contrary, then person X must concede. A counterexample like this has logical power. This is not just an ad hoc solution to an ad hoc problem: there is a whole class of similar problems with similar demonstrations. As practices, these are widely applicable, and can therefore be replicated.

One such practice is the differentiation of objects into partially overlapping sets by giving these objects one or more labels. The usability of this practice as a whole can be questioned – why differentiate the world into discrete terms in the first place? But remember from section 3.2 the way practices give meaning: given the practice, there is no questioning the meaning of a generalization and a refutation of a generalisation by way of counterexample. Every practice that gives meaning to generalizations, is a practice of counterexamples. Discrete differentiation has proven itself successful in the evolution of logic: modern logic still makes use of predicates.[168] In fact, Jc Beall et al. claim the modern model-theoretic account of logical consequence is based on developing "the intuitive idea of a counterexample […] into a theory of *models*."[169]

One reason for the flexible applicability of this practice, is that in principle, very little is required of the sets. When asked for the meaning of a term like virtue, everyone can consult their own set of experiences with virtuous things – practical situations in which the term gets meaning. As long as such a set has a clear label, they can be used in the practice of giving counterexamples. It is the collection of practical experiences with virtuous things that give virtue its meaning: first as a pre-theoretic notion, then later as a concept with clear relations to other concepts.

The logical practice of giving counterexamples has specific properties, that make generalisations and counterexamples to generalisations possible. The counterexamples give meaning to the notion of logical consequence, but also to other notions. I mention these notions here because the practice of giving counterexamples not only shows up in Aristotle's thoughts on logic, but also in those about the other notions.

- *Logical consequence*: each logical necessity rests on the refutation of generalisations by way of counterexample.

[168] Shapiro and Kouri Kissel, 'Classical Logic', sec. 2.1.
[169] Beall, Restall, and Sagi, 'Logical Consequence'.

- *Truth*:
    - Generalisations are true if counterexamples are impossible.
    - A counterexample is true if it can be given.
- *Entities* of truth: expressions about elements of sets.
- *Meaning*: the meaning of a term follows from a set of situations which the term applies to.
- *Proof*: true conclusions follow from true premises if
    - the premises do not allow a counterexample to the generalising conclusion, or if;
    - the premises do not allow generalisations to which the conclusion is a counterexample.

6.2.2 The practice before Aristotle

The practice of giving counterexamples to generalisations can be identified before Aristotle, in the work that is ascribed to Socrates (469 – 399 BC). I say ascribed, because Socrates himself has written nothing that we know of, and is used by Plato (429 – 347 BC) as a figure in his dialogues.[170] Where Plato's Socrates ends and the real Socrates begins is unclear. Although Aristotle was born after Socrates died, he did write about him, mostly following what had been said about him in Plato's dialogues.[171] This is unsurprising, because Aristotle was Plato's pupil.[172]

Socrates assumed the existence of truthful knowledge, but thought it was inaccessible by humans, who are marked by their ignorance.[173] In the context of my thesis, two philosophical methods that Socrates used are of importance – two reasoning practices:

1. The elenchus: this is a cross-examination meant to show the other that their assumed knowledge was not actual knowledge, by showing its inconsistency or incoherence.[174] These inconsistencies are derived using argumentation that the other must agree with – an indication that this method is based on a *logical* practice. This practice gave rise to the logic of the Stoics, which I get back to in section 6.3.
2. The dialectic method: this method is also a dialogue technique, but one with a more constructive purpose.[175] The goal is to systematically investigate one's own experiences, to reveal new knowledge. This method is based on the practice of giving counterexamples, as I will argue. Aristotle's conceptualization of this method leads him to his syllogistic logic.

---

[170] Debra Nails, 'Socrates', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2020 (Metaphysics Research Lab, Stanford University, 2020), sec. 2, https://plato.stanford.edu/archives/spr2020/entries/socrates/.

[171] James M. Ambury, 'Socrates', in *Internet Encylopedia of Philosophy*, sec. 1bv., accessed 21 February 2022, https://iep.utm.edu/socrates/.

[172] Christopher Shields, 'Aristotle', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2020 (Metaphysics Research Lab, Stanford University, 2020), sec. 1, https://plato.stanford.edu/archives/fall2020/entries/aristotle/.

[173] Ambury, 'Socrates', sec. 2bi.

[174] Ambury, sec. 3aii.

[175] Ambury, sec. 2c. As they do for anything ascribed to Socrates, historians debate about whether these two methods are distinct or whether they are two parts of one method. That the elenchus refutes and the dialectic constructs, seems to be acknowledged by most. So when I refer to the dialectic method, I am referring to the constructive part of Socrates' method(s).

How does the dialectic method work? I take my explanation of it from Gail Fine's book *On Ideas*.[176] Socrates starts by asking his interlocutor "what is F?" – where F is usually something of political or ethical nature, e.g. "what is virtue?" The interlocutor is expected to have a set of experiences with virtuous things or situations (Fs), which Socrates can use to give a counterexample. The interlocutor provides what is essentially a trial answer T, which can be wrong in two ways, according to Fine: it "can be both too narrow, if it fails to capture enough of the relevant instances, and too broad, if it captures some irrelevant ones."[177] From the perspective of my model, the trial answer T is based on pre-theoretic notions that, when put into words, provide a trial set of virtuous things and situations (Ts). The T is the desired F if the set of Ts and the set of Fs are the same. That is the case when both possible counterexamples are missing: there is no F that is not T, and no T that is not F.

The dialectic method can thus be seen as an inductive trial-and-error method with the counterexample as the error. The practice of giving counterexamples is used to establish a definition of a notion like virtue. The method assumes the existence of sets of experiences with practices that demonstrate this notion. The F that is found using this method, is worth nothing more than the set from which it has been derived inductively.

Plato admits as much in a dialogue he wrote, between Socrates and Meno.[178] Meno wonders how he can be expected to provide a trial answer T if he does not know what F is yet. How do we know what kind of answer is required? If we experience F in practice, how do we know it is F and not something else? Questions like these have to do with the genesis of sets of experiences.

In Plato's dialogue, Socrates first provides a summarization of Meno's paradox, which Fine formulates as follows:

> S1. For any x, one either knows, or does not know, x.
> S2. If one knows x, one cannot inquire into x.
> S3. If one does not know x, one cannot inquiry into x.
> S4. Therefore, for any x, one cannot inquiry into x.[179]

How Socrates – or rather Plato – solves the paradox, is not entirely clear.[180] Plato's most important point seems to be that the essence of F cannot be learnt from experiences. In Fine's words:

> A definition that simply listed the many Fs would be at best a nominal definition, telling us *that* certain things are F or are conventionally classified as Fs. It would not tell us *why* they are F, or whether our conventional classifications are correct. […] We can correctly believe that a particular action is just without knowing what justice is; but we cannot *know* whether anything is just without knowing what justice is.[181]

---

[176] Gail Fine, *On Ideas: Aristotle's Criticism of Plato's Theory of Forms* (Oxford, UNITED KINGDOM: Oxford University Press, Incorporated, 1995), 46–49, http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3053342.
[177] Fine, 47.
[178] Gail Fine, *The Possibility of Inquiry: Meno's Paradox From Socrates to Sextus* (Oxford: OUP Oxford, 2014), 7–9, https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=748993&site=ehost-live.
[179] Fine, 8.
[180] Fine, 10–12.
[181] Fine, *On Ideas*, 48–49.

When Socrates asks "what is F?", he is asking about Forms: real essences that explain why things are they way that they are.[182] Anything that is virtuous, is virtuous *because* it has the real essence of the Form of virtue. For Plato, Meno's paradox is reason to introduce his theory of recollection of innate ideas.[183] Plato claims that we have the ability to recollect what we knew of the Forms, because our soul retained this ability from when it was disembodied before we were born.[184] This theory of recollection is supposed to dissolve Meno's paradox, because on this account, inquiry is simply meant to activate recollection of the Forms.

Plato comes up with a whole hierarchy of perfect Forms, with at the top the One.[185] At the bottom of the hierarchy, we find a set of imperfect realizations of the Forms. The hierarchy is a way of discrete differentiation of the whole (the One) into parts.

6.2.3 Arguments for my thesis

In this section, I present some arguments that make plausible that Aristotle focused on sets – which he labels with universals – when developing his theory of logic. I argue that this put him on the track of the practice of giving counterexamples.

6.2.3.1 Aristotle as an empirical researcher

Aristotle had a strong affinity with empirical research. His father was a doctor.[186] Aristotle himself did research in marine biology, and he was well-versed in science in general.[187] Aristotle's school, the Lyceum, taught students in the methods of scientific inquiry.[188]

An argument for my thesis is that, given the facts above, we can conclude that Aristotle was familiar with the practice of discrete differentiation of the world into categories – and was thus also familiar with the practice of giving counterexamples. *Especially* as a biologist, he had experience categorizing individual organisms under a common denominator.[189]

6.2.3.2 Socrates' dialectic method as a starting point for Aristotle's logic

Aristotle describes predicates that apply to many things as *universals*.[190] Fine claims that Aristotle sees Socrates' Forms as universals: "[f]or since the form of piety is in everything that is pious, it is simultaneously in more than one thing at a time, and so it fits one of Aristotle's descriptions of

---

[182] Fine, 49.
[183] Glenn Rawson, 'Plato's Meno', in *Internet Encyclopedia of Philosophy*, sec. 2b., accessed 21 February 2022, https://iep.utm.edu/meno-2/.
[184] Richard Kraut, 'Plato', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2022 (Metaphysics Research Lab, Stanford University, 2022), sec. 1, https://plato.stanford.edu/archives/spr2022/entries/plato/.
[185] Kraut, sec. 1.
[186] Anton Dumitriu, *History of Logic*, ed. Duiliu Zamfirescu, Dinu Giurcăneanu, and Doina Doneaud, vol. 1 (Tunbridge Wells: Abacus Press, 1977), 142.
[187] Shields, 'Aristotle', sec. 1.
[188] Olaf Pedersen, *The First Universities: Studium Generale and the Origins of University Education in Europe* (Cambridge: University Press, 1997), 13.
[189] Pedersen, 13.
[190] S. Marc Cohen and C. D. C. Reeve, 'Aristotle's Metaphysics', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2021 (Metaphysics Research Lab, Stanford University, 2021), sec. 2, https://plato.stanford.edu/archives/win2021/entries/aristotle-metaphysics/.

universals."[191] She further quotes Aristotle as saying that "there are two things one might fairly ascribe to Socrates, inductive arguments and universal definitions, both of which are concerned with the starting-point of knowledge (*epistêmê*)."[192] So, Fine concludes: "Aristotle takes Socratic forms to be universals not just in the sense that they are or can be in more than one thing at a time but also in the sense that they are real essences, explanatory properties; for these are the sort of universals one needs to know in order to have knowledge."[193]

I argue that Aristotle takes Socrates' dialectic method as a starting point for his theory of logic. The F is not the essence, but what is *essentially* common to the set of Fs. Aristotle's "(dis)solution" to Meno's paradox is similar to Plato's. For Aristotle, foreknowledge of F comes not from recollection, but from a mental integration of repeated perception:

> From perception, then, as we say, memory arises. And from memory, when it occurs often of the same thing, experience arises; for memories that are many in number make up one experience. From experience, or (*ê*) from the whole universal that has settled in the soul (the one apart from the many, whatever is present as one and the same in all of them) there arises a principle of skill (if it is about what comes to be) or of knowledge (*epistêmê*) (if it is about what is).[194]

So, repeated perception is necessary for the universal to settle in the soul. Whether and how this settling can help perception discretely differentiate, say, cows from sheep, remains unclear: does this recognition of the universal act as a criterium for the establishing of sets of Fs, or does it help determine the universals after the sets of Fs have been established?[195] In the latter case, Aristotle does not solve Meno's paradox, because the paradox requires an explanation of how these sets of Fs can be established in the first place. In the first case, Aristotle's solution is similar to Plato's: they both require some special capacity to create the foreknowledge necessary to establish sets of Fs.

An argument for my thesis is that, though Aristotle might not provide a satisfactory solution to Meno's paradox, he does not seem to doubt the effectiveness of the dialectic method to provide universals: generalizations in the practice of the counterexample. This method is his starting point.

According to Fine, Aristotle did not intend to prove that Plato's metaphysics – the hierarchy of Forms – is based on false argumentation.[196] Aristotle provides an alternative metaphysics, one that better fits the dialectic method but maintains the hierarchical structure of Plato's metaphysics. At the top of the hierarchy of material things, Aristotle places *substance*, of which everything is made.[197] Underneath are the Forms – configurations of substance - which are meant to be understood as universals.

---

[191] Fine, *On Ideas*, 49–50.
[192] Aristotle, Metaphysics, 1078b; quoted in Fine, 50. Fine's translation.
[193] Fine, 50.
[194] Aristotle, Topics, 100a; quoted in Fine, *The Possibility of Inquiry*, 218. Fine's translation.
[195] Fine, 219.
[196] Fine, *On Ideas*, 28.
[197] Howard Robinson, 'Substance', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2021 (Metaphysics Research Lab, Stanford University, 2021), sec. 2.2, https://plato.stanford.edu/archives/fall2021/entries/substance/.

Taking Forms to be universals inevitably leads Aristotle to a different metaphysics, but the idea that everything is made up of one or more basic substances is not new, and was known to Aristotle.[198] However, speculating that everything is made up of, for example, water, does not straightforwardly lead to one way of categorizing the world. An argument for my thesis is that Aristotle changes Plato's metaphysics to bring it in line with the dialectic method, and therefore in line with the practice of giving counterexamples.

6.2.3.3 Aristotle's syllogistic logic is based on the practice of giving counterexamples

The dialectic method can also be used to investigate the relations between two concepts. Rather than comparing Ts and Fs, we can compare F1s and F2s: the sets that belong to the universals F1 and F2 and have been found with the dialectic method. There are four possible relations between the F1s and the F2s, described with the logical terms *all*, *no*, and *some*:

1. No F1 is F2 – true if the counterexample that is F1 and F2 is missing.
2. Some F1 is F2 – true if it can be demonstrated.[199]
3. All F1 is F2 – true if the counterexample that is F1 and not-F2 is missing.
4. All F2 is F1 – true if the counterexample that is F2 and not-F1 is missing.

From (3) and (4) together follows a fifth possible relation, where F1 = F2.

Aristotle's logic consists of logical arguments called syllogisms, that contain three terms A, B, and C.[200] A syllogism is made up out of two premises and one conclusion. One of the terms, called the middle term, is shared by both premises.[201] If B is the middle term, then a syllogism has the following form:

Premise 1: all/no/some A is B;
Premise 2: all/no/some B is C;

Conclusion: all/no/some A is C.

Negations like "not all" are also allowed.

If a premise is not a conclusion of some other syllogism, then its correctness depends on the dialectic method that provided the two terms used in the premise, and provided the correctness of the relation between them as described in the premise. If a premise *is* a conclusion of some other syllogism, then its correctness depends on the premises of that conclusion.

The meanings of the logical terms *all* and *no* are not straightforward in the case of empty sets. If A is empty, then both "all A is B" and "no A is B" can be said to be true, because without any As, both counterexamples would be missing: "A and not-B" and "A and B," respectively. But from Aristotle's perspective, a statement like "all A is B" when A is empty, would introduce a meaningless universal. That

---

[198] See Patricia Curd, 'Presocratic Philosophy', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2020 (Metaphysics Research Lab, Stanford University, 2020), sec. 2, https://plato.stanford.edu/archives/fall2020/entries/presocratics/ for an overview of some presocratic philosophers that were searching for 'first causes' like water.
[199] *Some* meaning "at least one."
[200] Robin Smith, 'Aristotle's Logic', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2020 (Metaphysics Research Lab, Stanford University, 2020), sec. 4.1, https://plato.stanford.edu/archives/fall2020/entries/aristotle-logic/.
[201] Smith, sec. 5.

is not to say that empty sets have no place. The intersection of the nonempty sets Round and Square is also empty (nothing is both round and square), but for sensible reasons. From the perspective of this intersection – in other words, *after* having differentiated the world into round and square things and nothing else – the only possible counterexamples are things that are round, square, or both. Then, only the generalizations “no Round is Square” and “no Square is Round” are true, because no counterexamples can be provided: the intersection is empty, there is nothing that is both round and square. For the generalisations “all Round is Square” and “all Square is Round,” there are nothing but counterexamples. There is a distinction between an empty set *without* label, and an empty set *with* label (like the set “Round and Square”). With deduction, all that counts is the label: the meanings of *all* and *no* depend on the sets you are starting with in the first place.

So, the logical power of the counterexample is not just useful to find universals and their relations, but can then also be used as premises to derive conclusions. Not all of the possible syllogisms are valid, however – for some syllogisms, the premises can be true while the conclusion is not. Aristotle found 14 valid syllogisms. I do not discuss all 14, but will show how for the syllogisms called Darapti, Barbara, Celarent, Darii, and Ferio, the logical power comes from the practice of giving counterexamples.[202] I chose Darapti because it is an example of an inference that is valid according to Aristotle’s logic, but not according to modern logic, so it will highlight the difference in the underlying practice. Barbara, Celarent, Darii, and Ferio are chosen because the other syllogisms can be derived from them using rules of conversion – which I discuss in section 6.2.3.4 – or by giving a counterexample to the whole syllogism.[203]

The syllogism Darapti looks as follows:

All A is B
All A is C

Some B is C

The conclusion is a counterexample to the generalization “no B is C” – in other words, either “some B is C” is true, “no B is C” is true, but not both. So, if Darapti is valid, then the premises should *refute* this generalization. That is possible if a counterexample to the generalization can be derived from the premises. Because the premises are generalizations themselves, the possible options they allow are determined by their counterexamples: elements that are A, but not B (for premise 1) and elements that are A, but not C (for premise 2). I use the following shorthand: $A\underline{B}*$ (counterexample to premise 1) and $A*\underline{C}$ (counterexample to premise 2), where $\underline{X}$ (underscored) indicates not-X, and * indicates either X or not-X.

Then, Darapti can be translated into the following:

All A is B	Not-($A\underline{B}*$)
All A is C	Not-($A*\underline{C}$)
-------------	--------------
Some B is C	Not-(no B is C)

[202] See Smith, sec. 5.4 for a list of all syllogisms.
[203] Smith, sec. 5.2.

The first premise, not-(A$\underline{B}$*), *refutes* the possibility of an element A$\underline{BC}$ as well as A$\underline{B}$C. The second premise, not-(A*$\underline{C}$), *also refutes* the possibility of an element A$\underline{BC}$, as well as AB$\underline{C}$. So, if something is A, it can only be both B and C, because ABC is the only remaining element.

This proof works if A is nonempty, because then we would have something that is A but also both B and C, which is what the conclusion claims. However, in case of an empty set A, Darapti would not be valid, but then a refutation would still make use of the practice of giving counterexamples. For a refutation, we would have to show that the premises *can* support "no B is C" (rather than "some B is C") – not that they *necessarily* support "no B is C". If A is empty, then, as I explained in the beginning of this section, it is just as right to say that "all A is B" as it is to say "no A is B." – remember the round/square example. But "no B is C" is not a counterexample to "no A is B" and no A is C". A counterexample to "no A is B" is something that is A and B (AB*). A counterexample to "no A is C" is something that is A and C (A*C). But these two counterexamples need not refer to the same thing, so it does not necessarily follow that it is something that is B and C.

Using the same shorthand as before, I translate Barbara in terms of counterexamples:

All A is B	Not-(A$\underline{B}$*)
All B is C	Not-(*B$\underline{C}$)
-------------	--------------
All A is C	Not-(A*$\underline{C}$)

The first premise, not-(A$\underline{B}$*), *refutes* the possibility of an element A$\underline{BC}$. The second premise, not-(*B$\underline{C}$), *refutes* the possibility of an element AB$\underline{C}$. That means that the possibility of A*$\underline{C}$ is also refuted: whether the element is B or not-B, either premise 1 or premise 2 refutes it. This is exactly the conclusion of Barbara: not-(A*$\underline{C}$), or "all A is C."

To be complete, below I provide the shorthand for the syllogisms Celarent, Darii, and Ferio.

Celarent

No B is A	Not-(AB*)	=> refutes ABC
All C is B	Not-(*$\underline{B}$C)	=> refutes A$\underline{B}$C
-------------	--------------	so, the possibility of A*C is refuted
No C is A	Not-(A*C)	

Darii

All B is A	Not-($\underline{A}$B*)	=> refutes $\underline{A}$BC
Some C is B	Not-(no C is B)	=> *BC not refuted
-------------	--------------	so, ABC refutes no C is A
Some C is A	Not-(no C is A)	

Ferio

No B is A	Not-(AB*)	=> refutes ABC
Some C is B	Not-(no C is B)	=> *BC not refuted
-------------	--------------	so, <u>A</u>BC refutes no C is <u>A</u>
Some C is <u>A</u>	Not-(no C is <u>A</u>)	

So, an argument is my thesis is that five syllogisms I discussed can be proven using only the practice of the counterexample. Together the rules of conversion that I discuss in the next section, this shows that Aristotle's whole syllogistic system is explainable in terms of the practice of giving counterexamples.

6.2.3.4 Aristotle's rules of conversion

Aristotle also used three rules of *conversion* to change syllogisms into other syllogisms:[204]

1. No A is B → no B is A.
2. Some A is B → some B is A.
3. All A is B → some B is A.

The first rule of conversion is evident: the counterexamples to "no A is B" and "no B is A" are the same, namely the element AB. Therefore, in the practice of giving counterexamples, they are equivalent. The same goes for the second rule of conversion: both "some A is B" and "some B is A" declare the possibility of an element AB. The third rule is more complicated. The counterexample to "all A is B" is the element A<u>B</u>. "Some B is A" declares the possibility of the element AB. This is only possible if the set A is nonempty, which is not refuted by "all A is B." The absence of A<u>B</u> does not prove the presence of AB, unless A is nonempty, which Aristotle assumes it to be.[205]

So, an argument for my thesis is that the validity of the syllogisms and their rules of conversion can be based on the practice of giving counterexamples. I have not discussed all 14 syllogisms, nor have I compared my proofs to Aristotle's proofs of the syllogisms. Further research is therefore warranted. Moreover, Aristotle did use counterexamples explicitly to prove the *invalidity* of the other possible syllogisms.[206] Furthermore, my thesis does not claim that the logicians in question should be aware that their concepts are based on certain practices, so if Aristotle is not explicit about the use of counterexamples in his proofs of the validity of the 14 syllogisms, that does not necessarily refute my thesis.

6.2.4 Other concepts connected to logical consequence

As I said in section 6.1, I now provide suggestions for the connections between the notion of logical consequence and the notions of *proof*, *truth*, *entities* of truth, and *meaning*.

The universals are true beliefs, if they are derived using the dialectic method – in other words, using the practice of giving counterexamples. The facts about the relations between universals as described by the premises of a syllogism, are also true beliefs. Aristotle distinguishes this truth from the *necessity* with which a conclusion follows from its premises: that is not a matter of true belief.[207] So, the '*inductive'*

[204] Smith, sec. 5.2.
[205] Smith, sec. 5.2.
[206] Smith, sec. 5.3.
[207] Fine, *The Possibility of Inquiry*, 206–7.

truth of the premises, which follows from the *meaning* of the terms, is distinct from the *necessary truth* of the conclusions, which follows from the form of the syllogism. The *entity* of truth is the link between two terms in the premises and conclusions of syllogisms. The valid syllogisms are the correct deductive *proofs*.

6.2.5 Conclusion

So, in conclusion, the evolution of Aristotle's logic has followed the following path:

1. First, the logic needs logical terms to reason deductively with. Induction, as based on the dialectic method, provides logical terms by holding concepts to be universals of sets.
2. Second, the logic needs a way to describe the relations between the logical terms. In Aristotle's syllogistic logic, relations between universals become relations between sets. This determines the form of the premises and conclusion of syllogism. These relations are the entities that can be true or false.
3. Third, the logic needs to describe the relations between these relations. Given three logical terms and two relations – which are bound by sharing the *middle* term – there is, in some cases, enough information to derive a *third* relation: the conclusion. In these cases, the truth of the conclusion rests on the absence of a counterexample, because the premises are said to be true if, using the dialectic method, *their* counterexamples cannot be provided.

6.3 Chrysippus

The limitations of the available resources are an even greater problem in the case of Chrysippus than they were in the case of Aristotle. Chrysippus became the third head of the Stoic school around 230 BC.[208] Of the first three heads of the Stoic school, only fragments of their work remain.[209] I rely on secondary and tertiary sources that have done the historical work for me, but the problem of a lack of information about the developments that let Chrysippus and the Stoics to their logical theory remains.

As in section 6.2, I first identify and discuss the relevant practice, in this case the practice of contradiction. Second, I explain how this practice existed before Chrysippus. Third, I provide some arguments to make plausible that it was this practice that led Chrysippus to his logical theory. Lastly, I provide a suggestion for the link between the notion of logical consequence, and the notions of *proof*, *truth*, *entities* of truth, and *meaning* – as explained in section 6.1 – and which I already preview in the next section.

6.3.1 The practice of contradiction

The idea behind the practice is that if we derive a contradiction using deductive reasoning, we have not reasoned correctly, or the premises already contained contradiction. Correct reasoning is defined by an enumeration of valid steps. No contradictions can be introduced nor eliminated during a *correct* deductive argument. So, the steps in an argument are valid if they do not allow for the addition or removal of a contradiction. A demonstration of a contradiction must therefore enumerate all those valid argumentation steps. In the following explanation of the practice of contradiction I take the proposition

---

[208] Jeremy Kirby, 'Chrysippus', in *Internet Encyclopedia of Philosophy*, sec. 1, accessed 28 February 2022, https://iep.utm.edu/chrysipp/.

[209] Dirk Baltzly, 'Stoicism', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2019 (Metaphysics Research Lab, Stanford University, 2019), sec. 1, https://plato.stanford.edu/archives/spr2019/entries/stoicism/.

as the entity of truth – which is what Chrysippus does – but the practice is valid for whatever entity of truth

What counts as a contradiction? The contradiction of a proposition does not necessarily have to be its negation, but can also be dependent on the meaning of the proposition and its contrary. For example, if the proposition is “X is taller than Y” and its contrary “X is shorter than Y”, then from the meanings of “taller than” and “shorter than” follows a contradiction. This becomes a *logical* contradiction (according to modern theory – see section 5.2) once we take as contrary not-B, meaning “not-(X is taller than Y)”. Not-B is different, as it includes the possibility that X and Y are the same length. However, the practice of the contradiction does not require the contrary to be a negation. In the rest of section 6.3 I write not-B for the contrary and leave the issue aside, unless it is required that the distinction be explicitly mentioned.

What, in practice, is a valid step from proposition A to proposition B? First, we must arrive at proposition A from a set of premises. Then, there must be a relation between A and B, along which the step is taken. If from A there is a relation to both B and the contrary of B (not-B), then it follows that the premises that led to A must contain a contradiction. So, from proposition A, there can only be a relation *either* to B, *or* to not-B, *or* to neither. In the latter case, the meanings of A and B do not allow a step from A to B or not-B.

The step from A to B or not-B is unidirectional, but not independent of the step from B or not-B to not-A. After the step from A to B, both the propositions A and B as well as their relations are valid. That means that the step from not-B to A is impossible, because not-B would already cause a contradiction with B. The step from B to not-A is invalid, because it would introduce a contradiction with A.

Assuming that the meanings of A and B are such that there is one connection between A and B, means that at proposition A, we are at a crossroads to either B or not-B. Just like with the practice of the counterexample in section 6.2.1, this is a matter of falsification: only when a step leads to a contradiction, it is false. The absence of a contradiction is temporary, as it can always be introduced later in the argument.

The practice of the contradiction requires that the propositions can be true or false in the same way as the unidirectional steps between them. In each basic argumentative step, two positions (A or not-A, B or not-B) and two connections (between e.g. A, and the possible conclusions B and not-B) are at stake. One premise concerns the validity of a connection, a second premise concerns the validity of position A, and a conclusion concerns the validity of the conclusion B or not-B.

This might all seem very similar to the practice of giving counterexamples. However, although a counterexample does contradict a generalization, the two propositions in a contradiction do not have to be a generalization and a counterexample. The practice of the contradiction does not require a set-theoretic interpretation to grant the counterexample the necessary logical power. This implies that the practice of the contradiction is more fundamental than the practice of giving counterexamples.

In the practice of the contradiction, logical consequence and its related notions get the following meanings. I mention these notions here because the practice of the contradiction not only shows up in Chrysippus’ thoughts on logic, but also in those about the other notions.

- *Logical consequence*: each logical necessity rests the on the avoidance of contradiction.

- *Truth*:
    - A collection of premises is true if they do not allow contradictions.
    - Conclusions are true if they can be derived from true premises.
- *Entities* of truth: propositions that can be true or false.
- *Meaning*: the meaning from a proposition follows from the thought that it is true.[210]
- *Proof*: a series of argumentative steps, starting with assuming the contrary of what you want to prove, and where each step is one of a collection of valid steps, all of which guarantee that no contradiction is introduced or eliminated.

6.3.2 The practice before Chrysippus

In section 6.2.2, I mentioned Socrates' constructive (dialectic) method as well as the elenchus, a method of refutation that reveals the inconsistency or incoherence in somebody's beliefs. I claimed that the dialectic method owes its success to the practice of giving counterexamples. The elenchus makes use of the practice of the contradiction. The elenchus uses argumentative steps to demonstrate that somebody's premises lead to a contradiction – which, as I claimed in section 6.3.1, does not necessarily have to consist of a proposition and its negation, but can also depend on the meaning of the two propositions involved in the contradiction.

Using this method, we can refute (sets of) premises, but that does not yet lead to better premises – at best, it increases the willingness to search for them. However, Zeno of Elea (495 – 430 BC) showed that deriving contradictions can also be used inductively, by accepting a contradiction as proof of the contrary of the premise that led to the contradiction. If the contradiction in question consists of a proposition and its negation, then this is called a *reductio ad impossibile*. Aristotle attributed the discovery of the use of this reductio in metaphysics to Zeno, although it is likely that Zeno himself got it from its use in Pythagorean mathematics.[211] The name *reductio ad absurdum* can be used for situations that are not strictly a *reductio ad impossibile*, because they allow for the contradiction to depend on the meaning of the respective propositions.[212] For both versions of the *reductio*, the goal is to accept a hypothesis by deriving a contradiction from its contrary. The subsequent rejection of the contrary of the hypothesis is seen as sufficient evidence for the acceptance of the hypothesis itself.

Clearly, then, the contradiction as criterium to distinguish true premises from false ones existed before Chrysippus.

6.3.3 Arguments for my thesis

In this section, I present some arguments that make plausible that Chrysippus focused on correct argumentative steps when developing his theory of logic. Although the roots of Stoic philosophy go back before Aristotle, Chrysippus was born after Aristotle. It is therefore likely he was aware of Aristotle's use of the dialectic method, as well as the differences between Plato's and Aristotle's underlying metaphysics and their approach of Meno's paradox.

---

[210] An untrue proposition has meaning for the person that holds it as true. That person might be confused, but their thoughts about the truth of the proposition explain its meaning, even if it is not actually true. Meaning follows from taking something to be true.

[211] William Carvert Kneale and Martha Kneale, *The Development of Logic*, 6th edition (Oxford: Clarendon Press, 1975), 7.

[212] Kneale and Kneale, 9.

6.3.3.1 Chrysippus' focus on determinism

Stoicism is first and foremost a naturalistic philosophy. Logic, physics, and ethics are all related because even when the Stoics mention an abstract concept like 'soul,' they are referring to physical entities.[213] These physical entities are known to us by use of a rational principle that is part of the universe.[214]

In Stoicism, logic describes the workings of nature, which the virtuous human should follow.[215] For Chrysippus, fate is related to causation.[216] For everything, there must be sufficient cause or reason. Chrysippus' determinism is universal, but also complex. Having to relate logic, physics, *and* ethics, Chrysippus is presented with the problem of having to reconcile fate with human choice and responsibility.[217]

An argument for my thesis is that despite this problem, Chrysippus never doubts the strong connection between logic, physics, and ethics. His determinism makes his commitment to the practice of the contradiction self-evident: if everything from physics to ethics is describable using the same rational principle that also underlies logic, then contradiction becomes an important tool to understand the world.

6.3.3.2 Acceptance of the principle of bivalence

Using contradictions inductively, as in the *reductio*, is only possible given the assumption that the principle of bivalence applies to all argumentative steps: we are either at proposition A or not-A (but not at both), and from each of these propositions either B or not-B is derivable (but not both). Because propositions *and* the relations between propositions can both be premises, the principle of bivalence must also apply to each possible relation. It is in principle possible for *both* relations (between A and B and between A and not-B) to be absent, but a citation attributed to Chrysippus by Cicero shows that Chrysippus rejects this option: "If uncaused motion exists, it will not be the case that every proposition (termed by the logicians an *axioma*) is either true or false, for a thing not possessing efficient causes will be neither true nor false; but every proposition is either true or false; therefore uncaused motion does not exist."[218]

To prove uncaused motion does not exist, Chrysippus assumed that every proposition is either true or false. So, the principle of the bivalence is key to this argument. Chrysippus did not choose for the practice of the contradiction and therefore had to accept the principle of bivalence too (or vice versa): for him, they are one and the same. So, an argument for my thesis is that from his acceptance of the principle of bivalence, we can assume that Chrysippus accepts the practice of the contradiction. As I explained in section 6.3.1, this practice can be used inductively (per *reductio*), just like the practice of giving counterexamples can be used inductively using Aristotle's constructive dialectic method.

---

[213] Massimo Pigliucci, 'Stoicism', in *Internet Encyclopedia of Philosophy*, sec. 2, accessed 28 February 2022, https://iep.utm.edu/stoicism/.
[214] Pigliucci, sec. 2.
[215] Baltzly, 'Stoicism', sec. 5.
[216] Baltzly, sec. 3.
[217] Baltzly, sec. 3. Chrysippus' solution is to distinguish different kinds of causes, but that need not concern us here.
[218] Cicero, 'On Fate: Chapter X', in *On the Orator: Book 3*, trans. Harris Rackham, Loeb Classical Library (Cambridge, MA: Harvard University Press, 1942), paras 20–21, https://www.loebclassics.com/view/marcus_tullius_cicero-de_fato/1942/pb_LCL349.217.xml.

Chrysippus equates causal relations to logical relations: an uncaused effect X can be neither true nor false. If X is uncaused, no premises are possible from which we can derive X or not-X by following valid argumentative steps. A non-causal world is unthinkable for a Stoic like Chrysippus, so the meaning of X must lie in its relation to other causes and effects. Each X either is or is not partly the cause of each Y – directly or indirectly. Then, for Chryssipus, the possibility of a proposition A without relation to either B or not-B is ruled out. Otherwise, we would have a causal network that includes A but which is not part of another separate causal network that includes B or not-B.

An argument for my thesis is that Chrysippus does not only consider the principle of bivalence to apply to propositions and relations between propositions, but also to the crossroad to either B or not-B from each proposition A. From each proposition A, there is always one relation to either B or not-B, regardless of the meanings of A and B. Therefore, the practice of the contradiction can be the only criterium to stop the argumentation at any given argumentative step.

6.3.3.3 Chrysippus' rejection of the universality of Aristotle's universals

In section 6.3.3.2, I said that both Chrysippus and Aristotle apply their respective practices inductively – Chrysippus through use of the *reductio*, Aristotle through use of his dialectic method. In this section, I argue that Chrysippus rejects the universality of Aristotle's universals, and therefore the inductive power of Aristotle's dialectic method. However, Chrysippus still needs an answer to Meno's paradox.[219] I claim that Chrysippus gets the required inductive power from the practice of the contradiction, rather than the practice of giving counterexamples.

An important idea is that of *prolepses*. According to Fine, it is a difficult concept to define.[220] The Stoics considered them the answer to Meno's paradox.[221] How do we, inductively, inquire into unknown subjects? The Stoics took prolepses to be necessary for inquiry. They see them as "true propositions about basic features of reality; and they think we acquire prolepses naturally, without teaching or learning."[222] Together with sense-perception, Chrysippus considered prolepses to be a criterium for truth.[223]

According to the Stoics, concepts are not a requirement for experience.[224] True propositions based on initially nonconceptual experience are possible through what they called *apprehension* – something stronger than opinion but less stable than actual knowledge, though on the way to actual knowledge.[225] The Stoics thought that being rational means having a set of concepts and prolepses, having *conceptualized* mental representations of the world.[226] This conceptualized mental representation – knowledge – is possible because of our prolepses.

---

[219] See section 6.2.2.
[220] Fine, *The Possibility of Inquiry*, 228.
[221] Fine, 297.
[222] Fine, 297.
[223] Fine, 263–64.
[224] Fine, 259.
[225] Pigliucci, 'Stoicism', sec. 2a.
[226] Fine, *The Possibility of Inquiry*, 260.

Prolepses identify not essences, but thoughts.[227] In fact, the Stoics thought (Aristotle's) universals are "figments of the mind."[228] They were nominalist in that sense.[229] The prolepses form the basis for all concepts, not by being definitions themselves, but by being (expressed by) "outline accounts."[230] The outline account is the starting point for definitions. Rather than treating words like 'man' and 'mortal' as referring to universals, Chrysippus analysed definitions as conditional propositions: e.g., if something is a man, then it is mortal.[231] Defining is a process of constructing relations. Prolepses are the certain starting point to do this from: if X has such outline account, then X also has this more essential property.

So, the prolepses, usable without further definition, make nominalism sufficient to answer Meno's paradox. Aristotle's universals are not acceptable to Chrysippus, because Aristotle assumes that universals must be shared by multiple things – for example, all particular men share the essence of being mortal.[232] But the Stoics considered universals to be figments of the mind.

The prolepses can be considered as true propositions, that can be nonconceptually taken from experience. So, an argument for my thesis is that the process of definition, with the prolepses as starting point uses the practice of the contradiction, where the consistency and coherence of the prolepses is preserved. Thanks to the prolepses, proofs that make use of the *reductio* are possible. Resting on the prolepses, the system is not waterproof, but that is not a problem for the Stoics: they thought of humans knowing anything as the exception rather than the rule.[233]

6.3.3.4 Chrysippus' five indemonstrable valid arguments

In this section, I show that Chrysippus' five indemonstrable valid arguments can be understood as a complete list of argumentative steps, all of which neither add nor eliminate contradiction. As I explained in section 6.3.1, this is what I take the practice of the contradiction to entail.

The Stoics use the logical operators to create composite propositions, which can then take the place of singular propositions in argumentative steps. The question, then, is which elementary propositions there are: the ones that cannot be reduced into smaller steps, and that do not introduce nor eliminate contradiction. These axiomatic arguments are called indemonstrables.[234]

Chrysippus refers to the two propositions in an argumentative step with ordinal numbers (first, second) rather than letters.[235] For readability, I will use the numbers 1 and 2. His axiomatic arguments, like Aristotle's syllogisms, use two premises and a conclusion. The first premise is a relation between 1 and 2, and the second premise is either 1 or not-2.

---

[227] Fine, 260.
[228] Baltzly, 'Stoicism', sec. 3.
[229] Baltzly, sec. 3.
[230] Fine, *The Possibility of Inquiry*, 267.
[231] Henry Dyson, *Prolepsis and Ennoia in the Early Stoa* (Berlin/Boston, Germany: De Gruyter, Inc., 2009), 101, http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=476070.
[232] See section 6.2.3.2.
[233] Baltzly, 'Stoicism', sec. 4.
[234] Susanne Bobzien, 'Ancient Logic', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2020 (Metaphysics Research Lab, Stanford University, 2020), sec. 5.4, https://plato.stanford.edu/archives/sum2020/entries/logic-ancient/.
[235] Kneale and Kneale, *The Development of Logic*, 163.

Each of the following five indemonstrables do not introduce nor eliminate contradiction. For each step, the following applies:[236]

- If a relation between 1 and 2 is valid, then the relation between 2 and not-1 must be invalid, because otherwise, the step would introduce a contradiction.
- The principle of bivalence applies to all propositions and relations, as well as to the two possible relations between one proposition and another or its contrary – of these two relations, one and only one is true.

A. If 1 then 2; but 1; therefore 2.

This indemonstrable is rightly called an indemonstrable, because the proof should make use of this step. To prove this indemonstrable, we would have to start by assuming the contrary of the conclusion, so not-2. Then, from the second premise follows 1. Given 1, from the first premise follows 2, but only if we use this same indemonstrable, which is what we set out to prove.

The possibility not-2 is incompatible with 1. Nothing is said about the possibility of not-1, which can have a relation with either 2 or not-2. But a relation from 2 to not-1 should be ruled out, otherwise this step would not be valid, as we would have a contradiction with the second premise: 1.

B. If 1 then 2; but not-2; therefore not-1.

Again, we assume the contrary of the conclusion, so 1. From the first premise, using indemonstrable A, 2 follows. But this is in contradiction with the second premise, namely not-2. So, the practice of the contradiction, together with indemonstrable A, can prove the validity of B.

The possibility of 1 is incompatible with not-2. There is a parallel with indemonstrable A: B can be written in the same form as A, if we replace 1 with not-2, and 2 with not-1 (and vice versa). Then we get **"if not-2 then not-1; but not-2, therefore not-1."** Nothing is said about the possibility of 2, which can have a relation with either 1 or not-1. But a relation from not-1 to 2 should be ruled out, otherwise this step would not be valid, as we would have a contradiction with the second premise: not-2.

C. Not both 1 and 2; but 1; therefore not-2.

Again, we assume the contrary of the conclusion, so 2. Together with the second premise follows 1 and 2. But this is in contradiction with the first premise: not both 1 and 2.

The first premise also includes information that is unnecessary to conclude not-2, namely "if 2, then not-1." We can strip the indemonstrable from this unnecessary information, by again rewriting it in the same form as A, this time by replacing 2 with not-2. Then we get **"if 1 then not-2; but 1; therefore not-2."** Nothing is said about the possibility of not-1, which can have a relation with either 2 or not-2. But a relation between 2 and not-1 should be ruled out, otherwise this step would not be valid, as we would have a contradiction with the second premise: 1.

In its rewritten form, C is self-referential in the same way as A is. C is truly different from A in this way, unlike B, which can be proven using A.

[236] This is a summary of section 6.3.1.

D. Either 1 or 2; but 1; therefore not-2.

Again, we assume the contrary of the conclusion, so 2. Together with the second premise follows 1 and 2. But this is in contradiction with the first premise: either 1 or 2 (but not both).

The first premise also includes the unnecessary information "if 2, then not-1," "if not-1, then 2," and "if 1, then not-2." We can strip the indemonstrable from this unnecessary information, by again rewriting it in the same form as A, and again replacing 2 with not-2, as in C. Then we get **"if 1 then not-2; but 1; therefore not-2."** Nothing is said about the possibility of not-1, which can have a relation with either 2 or not-2. But a relation between 2 and not-1 should be ruled out, otherwise this step would not be valid, as we would have a contradiction with the second premise: 1.

C and D are not the same if the unnecessary information is not stripped. If we were to add the premise "if not-1, then not-2," then this would contradict the first premise of D, but not any premises of C. Nevertheless, this changes nothing about the conclusion.

E. Either 1 or 2; but not-2; therefore 1.

Again, we assume the contrary of the conclusion, so not-1. From the principle of bivalence and the first premise follows 2. But this is in contradiction with the second premise, namely not-2.

The first premise also includes the unnecessary information "if 2, then not-1," "if not-1, then 2," and "if 2, then not-1." We can strip the indemonstrable from this unnecessary information, by again rewriting it in the same form as A, this time replacing 1 with not-2, and 2 with 1. Then we get **"if not-2 then 1; but not-2; therefore 1."** Nothing is said about the possibility of 2, which can have a relation with either 1 or not-1. But a relation between 1 and 2 should be ruled out, otherwise this step would not be valid, as we would have a contradiction with the second premise: not-2.

In its rewritten form, E is self-referential in the same way as A and C are.

With these indemonstrables, Chrysippus is searching for those elementary argumentative steps that do not allow the introduction nor elimination of contradiction. He formulates these in logical terms, as they would occur in an argument. For B, C, D, and E, this leads to premises that include more information than necessary to draw the conclusion. If we strip these arguments from this information, we are left with three axiomatic arguments, which all have the same general form as A: A itself, C rewritten as A, and E rewritten as A. B is derivable from A, and D is the same as C.

This list of the three axiomatic arguments of the same form is not complete, as there are four pairs of arguments. For each pair, the right one is derivable from the left one (or vice versa). I have already shown this for the first pair.

If 1 then 2; but 1; therefore 2 (A)	If not-2 then not-1; but not-2, therefore not-1 (B)
If 1 then not-2; but 1; therefore not-2 (C and D)	If 2 then not-1; but 2; therefore not-1
If not-2 then 1; but not-2; therefore 1 (E)	If not-1 then 2; but not-1; therefore 2
If 2 then 1; but 2; therefore 1.	If not-1 then not-2; but not-1; therefore not-2

Of the five indemonstrables that Chrysippus mentions, only four are unique of the eight arguments in the above schema. But these other arguments follow from the information we stripped by translating the indemonstrables into the same general form of A. So, an argument for my thesis is that Chrysippus sums up all possible forms of elementary steps that do not introduce nor eliminate contradiction, which is what the practice of the contradiction requires.

6.3.3.5 Chrysippus' themata (inference rules)

Aside from the axiomatic indemonstrables, the Stoics also had four inference rules called *themata*.[237] If an argument is an indemonstrable itself, or reducible to one by use of these themata, then it is valid. In this section, I argue that none of these themata add or eliminate contradiction, which is what I take the practice of contradiction to entail. Only the first and third themata are known. Reconstructions of the other two themata have been attempted, but for my thesis I limit myself to the themata that are extant.

The first themata reads: if "A; and B; therefore C" follows, then also "A; but not-C; therefore not-B" and "B; but not-C; therefore not-A".[238]

In section 6.3.3.4, I derived B from A. This proof made use of the first themata, by replacing the first argument in the themata with "if 1 then 2; but 1; therefore 2" (indemonstrable A). Then, according to the first themata, also the conclusions "if 1 then 2; but not-2; therefore not-1" (indemonstrable B) and "1; but not-2; therefore not-(if 1 then 2)" should follow. So, the themata is more universal, because it can not only derive indemonstrable b, but also an argument of the form "1; but not-2; therefore not-(if 1 then 2)".

We should then prove that this second argument follows from indemonstrable A using the practice of the contradiction. So, assume the contrary of the conclusion of this second argument, "if 1 then 2". But from the first premise follows 1. Using indemonstrable A then follows 2. But this is in contradiction with the second premise, namely not-2.

The third themata reads: if "A; and B; therefore C" follows, and "C; and D; therefore E" follows, then also "A; and B; and D; therefore E" follows.[239]

To prove this using the practice of the contradiction, assume the conclusion is false, so it follows that not-E, A, B, and D. From "C; and D; therefore E" follows that C or D must be false. But D is true, so C must be false. But then from "A; and B; therefore C" follows that A or B must be false. This is in contradiction with the assumption that A and B are both true.

So, an argument for my thesis is that Chrysippus' themata do not introduce nor eliminate contradiction, which is what the practice of the contradiction requires. This, as I have shown in section 6.3.3.4, is also the case for his indemonstrable arguments. Neither make use of any truth-tables, which is relevant for the debate on whether Chrysippus adhered to a truth-functional definition of the "if… then…" conditional.[240] I have shown that at the very least, he did not *need* to.

[237] Bobzien, 'Ancient Logic', sec. 5.4.
[238] Bobzien, sec. 5.4. Note that these three arguments have the same syllogistic form of two premises and a conclusion as the indemonstrables.
[239] Bobzien, sec. 5.4.
[240] See e.g. Joseph M. Bochenski, *A History of Formal Logic*, ed. Ivo Thomas (Notre Dame: University of Notre Dame Press, 1961), 119.

6.3.4 Other concepts connected to logical consequence

As I said in section 6.1, I now provide suggestions for the connections between the notion of logical consequence and the notions of *proof*, *truth*, *entities* of truth, and *meaning*.

Propositions are true beliefs, if they are established using *reductio* – in other words, using the practice of the contradiction. Concepts only have *meaning* within the context of a proposition, which as a whole also has meaning. Chrysippus distinguishes this from the necessity of the conclusion given the premises. That is not a matter of true belief. So, the '*inductive' truth* of the premises, which follows from the *meaning* of the propositions, is distinct from the *necessary truth* of the conclusions, which follow from the form of the argument. The *entity* of truth is the proposition. The valid arguments are the full list of correct deductive *proofs*: the indemonstrables.

6.3.5 Conclusion

So, in conclusion, the evolution of Chrysippus' logic has followed the following path:

1. First, the logic needs logical propositions to reason deductively with. Chrysippus' induction is based on proofs of propositions using *reductio ad absurdum* or *impossibile*, with the prolepses as a starting point.
2. Second, the logic needs a way to describe the relations between the logical propositions. For Chrysippus, these relations are also propositions. The relations between the concepts in a proposition and the meaning of the proposition as a whole, determine the meanings of the concepts. Propositions can be true or false. Premises and conclusions in an argument are propositions. Proof by *reductio* is used to ascribe truth to premises.
3. Third, the logic needs to describe the relations between these relations. Given a proposition A and its relation with another proposition which can be either B or not-B, there is in some cases enough information to decide whether this other proposition is B or not-B. In these cases, the truth of the conclusion depends on the absence of contradiction, because the premises – thanks to proofs by *reductio* – can be called true because they do not lead to contradiction.

Conclusion

I started this thesis by wondering how to explain change in knowledge. I began by considering evolutionary epistemologies, which have linked change in theories to change in species. On one simplified view, both theories and species change because they adapt to their environment. As I explained in chapter 2, however, this view has limitations. An alternative, non-adaptationist approach has been put forward that, when it comes to change, rather than imparting all causal significance to the environment, grants individual organisms the agency to play their own causal role in their evolution. In biology, this view is called the Extended Evolutionary Synthesis. I ended chapter 2 with an explanation of this view and how it might inform evolutionary epistemology.

Even when abandoning the adaptationist approach, however, we are still left with the problem that concepts and organisms are not the same entities, and those properties that apply to organisms, cannot without abandon be attributed to concepts. Evolutionary epistemologies have therefore often focused on the *analogies* between conceptual change and biological change. This metaphorical approach creates its own problems, which I discussed chapter 3. I concluded that we need a way to relate conceptual change to evolution that does not focus on the similarities, but rather *extends* evolutionary theory to include concepts. To do so, I connected conceptual change to change in practices: intentional, replicable behaviour of agents. For this, I introduced some tools taken from the later Ludwig Wittgenstein's ideas on meaning and language-games. The main takeaway is that the bedrock of our concepts exists not in some other, foundational concepts, but in the practices in which we use them. Practices invoke certain pre-theoretic notions which can be conceptualized. I ended chapter 3 with some remarks on the self-referential nature of my model and its use of the notion of a *concept*.

In chapter 4, I proposed a model that explains the interdependencies between conceptual change, change in practices, and change in organisms. It is built up from three notions: of *organism*, *agency*, and *concept*. These three notions give rise to three distinctions: between the inorganic and the organic, between passive attributes and active practices, and between demonstrations and concepts. This results in a model of three evolutions: of organisms, practices, and concepts. The model describes the interdependencies between these evolutions: concepts need practices (based on Wittgensteinian theory) and practices need organisms, but practices can also affect the success of organisms in their evolution (based on the Extended Evolutionary Synthesis), and concepts can also affect the success of practices. I then ended chapter 4 with some remarks on the issue of defining *correctness* of knowledge in light of my model.

In chapter 5, I introduced knowledge of logic as a case study for my model, and discussed Stewart Shapiro's and Graham Priest's ideas on the legitimacy of logical theories. I claimed that logic is about the notion of logical consequence. I then discussed six pre-theoretic notions of logical consequence, as outlined by Shapiro. In the last two sections, I used the views on legitimacy of Shapiro and Priest, respectively, to highlight potential misunderstandings of my model.

In chapter 6, I applied my model to a case study in logic, focusing on the notion of *logical consequence*. I discuss two conceptualizations of logic and logical consequence in ancient history: that of Aristotle and that of Chrysippus. For Aristotle, I argued that it was the practice of giving counterexamples that gave rise to his syllogistic logic. For Chrysippus, I argued that it was the practice of the contradiction that gave rise to his Stoic logic.

Closing thoughts

Further research on Frege

Both for Aristotle and Chrysippus, their respective practices are first employed for inductive purposes. For both, their respective inductive methods form the basis for a deductive method, in which conclusions follow out of *necessity*. How does this work? The inductive methods create premises that are true in some way, and this truth is meant to be preserved using deduction. Whatever form of truth the premises had, the conclusion must have too. In this sense, induction and deduction in the time of Aristotle and Chrysippus shared the same core: truth at absence of counterexample or contradiction. Deduction also has the same form for both Aristotle and Chrysippus: two premises, one conclusion.

Potential further research should involve Gottlob Frege (1848 – 1925) and consider his attitude towards the two practices. It seems that he uses a notion of a function to integrate the practice of giving counterexamples with the practice of the contradiction. He combines the generalization of Aristotle with the conditional of Chrysippus: Aristotle's "all P is Q" becomes "for all x, P(x) → Q(x)", where "→" is Chrysippus' conditional, in adapted form – meaning "P(x) → Q(x)" is only false if P(x) is true and Q(x) is false. That means that unlike in Aristotle's logic, "all P is Q" is also true when P is empty – in other words, when the function P applies to no x. A further consequence is that "not-Q(x) → Q(x)" is true if Q(x) is true or false – in other words, it is *always* true. This is difficult to analyse in terms of the practices of giving counterexamples or of the contradiction.

A second problem is that one would now expect that logical operators like AND could also be written as functions. However, they do not accept x as input, but only functions that themselves accept x as input (P(x) and (Q(x)) – AND(P(x), Q(x)) is possible, but not AND(x, y). In that sense, they belong to second-order logic, where variables can also vary over the domain of functions. This too is difficult to analyse in terms of the practices of giving counterexamples or of the contradiction.

The issue is not that Frege does away with the logical force of giving a counterexample or showing that the contrary leads to a contradiction. I think these two remain the core of deduction. The problem is that Frege adopts a new practice of axiomatising, which introduces a set of well-chosen related concepts, that seem sufficient to define an absolute certain practice of deduction. Logic then seems grounded in these axioms, which create practices rather than the other way around. Frege's axiomatic logic is designed to be a logic for axiomatic mathematics. The idea of grounding in axioms offers freedom: one could design axioms for special logics for certain purposes or certain domains of application. This freedom raises the question of how we know which set of axioms is better for which purpose of domain. Axiomatising practices is in itself a practice too, with 'axiom' as its core concept. This complicates an analysis of the Frege's logic in terms of the practices of giving counterexamples or of the contradiction.

Limitations of my model

I think the biggest result of my thesis is that it offers an alternative way to explain correctness. Figuring out which theory is more correct is easy enough when everyone agrees on what counts as evidence, but things can become more complicated when other criteria for correctness like simplicity are brought in. Rather than a focus on which of these are the right criteria, and how much weight each should have when deciding on competing theories, my model suggests that we should consider what practices we are operating in. Two scholars who claim to disagree on a certain concept might not really be in disagreement if it turns out they are using it in the context of a different practice. If the answer would be that the scholars are speaking of the concepts in absolute terms, then, according to my model, they would have no meaning, and there would not really be any disagreement either.

I end the thesis by pointing out some limitations of my model. As we have seen in chapter 2, evolutionary theories generally aim to explain *variation*, *selection*, and *retention* of traits in organisms. Evolutionary epistemologies tend to do the same for theories. As I discussed in section 3.1, Fracchia and Lewontin criticize this metaphorical approach. I claimed to avoid their criticism because my model is not an analogy but an extension of evolutionary theory, one that adds change of practices and concepts to change of organisms. That is why I refrain from calling it an evolutionary epistemology. However, that does not mean that variation, selection, and retention do not require an explanation for practices and concepts, just like they do for organisms. Because of the scope of my thesis, I must forego a precise explanation of these notions, but below I provide some ideas on what I think is lacking, and how to proceed in potential further research.

I claimed that organisms, as agents, are not just the effect, but also the cause of the variation, selection, and retention of their own traits. The same claim cannot intuitively be made of practices and concepts. My model does explain some parts of the selection and retention – which has to do with their grounding in practices – and the selection and retention of practices – which has to do with their grounding in attributes. In section 2.3, I have addressed the replication of practices and concepts, which happens through e.g. parent-offspring or teacher-student relations, and is sometimes called cultural or horizontal replication. However, intuitively it seems that practices and concepts do not *vary* themselves: *we* vary them. I have not provided an explanation of where this variation comes from beyond that it comes from our agency as organisms. Perhaps a similar claim can be made not only of practices and concepts, but also of attributes, but that requires a way to explain in agential terms the important role of genetics in evolution.

I think the issue is, fittingly, one of grounding. On one hand, my model does consist of a loop. Concepts need practices, and practices need organisms, but concepts also affect the survival of practices, and practices affect the survival of organisms: there is reciprocal causation between the three evolutions. On the other hand, by implicitly taking agency to be some sort of attribute of organisms, I am forced to explain the evolution of practices and concepts in terms of the agency of organisms. A more synergistic model would have to employ a more broadly applicable definition of agency.

A related issue is that of *levels of selection*. When we are asked to explain variation in evolution, this generally refers to *intra*variation, the variation that is observable within a population or species. This is the sort of variation that, after long evolutionary timelines, gives rise to the *inter*variation we observe *between* species. Similarly, when I speak of the practices of Aristotle or Chrysippus, I am not referring to one specific practice that they partook in, distinct from other practices. Rather, I refer to a group of practices that, like a group of organisms, is self-similar enough to be called one practice, or one species. For organisms, this is done on the basis of some account of variation, which my model lacks.

Another matter I have not addressed is the matter of progress in knowledge. If, as I claimed in section 4.4, taking my model seriously means correctness should not be seen in an absolute sense anymore, then is there still a way to say that our knowledge gets better over time? Presumably, an answer to this question would be a similar answer to the question of whether species can be said to become “better” over time. I have not provided a precise account of what it means for an organism, practice, or concept to be better or more correct, except to say that the success of each rests on its selection and replication. The success of a concept of a hammer, which I discussed in section 4.4, is relatively easy to explain in relation the intention behind the practice. So is the success of a practice like eating, because it serves to alleviate experiences of hunger. Things are not so straightforward with more complicated practices and concepts like in the case of logic.

Such an answer is not likely to appease critics that would claim that my model does not explain progress of knowledge. Indeed, it does not. However, what I *can* say is that such criticism must first assume there is such a thing as progress in knowledge in the first place. One can only speak of progress with reference to some criterium. As the role and even existence of absolute correctness becomes muddied when taking my model seriously, asking why there is progress in knowledge amounts to begging the question. Nevertheless, my model should at least aim to explain why we seem to *experience* progress in knowledge. I leave that undertaking for the future.

Bibliography

Ambury, James M. 'Socrates'. In *Internet Encyclopedia of Philosophy*. Accessed 21 February 2022. https://iep.utm.edu/socrates/.

Baedke, Jan, Alejandro Fábregas-Tejeda, and Guido I. Prieto. 'Unknotting Reciprocal Causation between Organism and Environment'. *Biology & Philosophy* 36, no. 5 (October 2021): 1–29. https://doi.org/10.1007/s10539-021-09815-0.

Baedke, Jan, Alejandro Fábregas-Tejeda, and Francisco Vergara-Silva. 'Does the Extended Evolutionary Synthesis Entail Extended Explanatory Power?' *Biology and Philosophy* 35, no. 1 (January 2020): 1–22. https://doi.org/10.1007/s10539-020-9736-5.

Baltzly, Dirk. 'Stoicism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/spr2019/entries/stoicism/.

Beall, Jc, Greg Restall, and Gil Sagi. 'Logical Consequence'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/spr2019/entries/logical-consequence/.

Ben-Yami, Hanoch. 'Vagueness and Family Resemblance'. In *A Companion to Wittgenstein*, edited by Hans-Johann Glock and John Hyman, 407–19. Chichester: John Wiley & Sons, 2017.

Bobzien, Susanne. 'Ancient Logic'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/sum2020/entries/logic-ancient/.

Bochenski, Joseph M. *A History of Formal Logic*. Edited by Ivo Thomas. Notre Dame: University of Notre Dame Press, 1961.

Boghossian, Paul. 'Knowledge of Logic'. In *New Essays on the a Priori*, edited by Paul Boghossian and Christopher Peacocke, 229–54. Oxford, UNITED KINGDOM: Oxford University Press, Incorporated, 2001. http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3052772.

Bradie, Michael. 'Assessing Evolutionary Epistemology'. *Biology and Philosophy* 1, no. 4 (1986): 401–59.

Cicero. 'On Fate: Chapter X'. In *On the Orator: Book 3*, translated by Harris Rackham. Loeb Classical Library. Cambridge, MA: Harvard University Press, 1942. https://www.loebclassics.com/view/marcus_tullius_cicero-de_fato/1942/pb_LCL349.217.xml.

Cohen, S. Marc, and C. D. C. Reeve. 'Aristotle's Metaphysics'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/win2021/entries/aristotle-metaphysics/.

Curd, Patricia. ‘Presocratic Philosophy’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/fall2020/entries/presocratics/.

Dumitriu, Anton. *History of Logic*. Edited by Duiliu Zamfirescu, Dinu Giurcăneanu, and Doina Doneaud. Vol. 1. Tunbridge Wells: Abacus Press, 1977.

Dyson, Henry. *Prolepsis and Ennoia in the Early Stoa*. Berlin/Boston, Germany: De Gruyter, Inc., 2009. http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=476070.

Ebbs, Gary. ‘Rules and Rule-Following’. In *A Companion to Wittgenstein*, edited by Hans-Johann Glock and John Hyman, 390–406. Chichester: John Wiley & Sons, 2017.

Fine, Gail. *On Ideas: Aristotle’s Criticism of Plato’s Theory of Forms*. Oxford, UNITED KINGDOM: Oxford University Press, Incorporated, 1995. http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3053342.

———. *The Possibility of Inquiry: Meno’s Paradox From Socrates to Sextus*. Oxford: OUP Oxford, 2014. https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=748993&site=ehost-live.

Fracchia, Joseph, and Richard Charles Lewontin. ‘The Price of Metaphor’. *History and Theory* 44, no. 1 (February 2005): 14–29.

Gibson, Edward, Richard Futrell, Julian Jara-Ettinger, Kyle Mahowald, Leon Bergen, Sivalogeswaran Ratnasingam, Mitchell Gibson, Steven T. Piantadosi, and Bevil R. Conway. ‘Color Naming across Languages Reflects Color Use’. *Proceedings of the National Academy of Sciences* 114, no. 40 (3 October 2017): 10785–90. https://doi.org/10.1073/pnas.1619666114.

Glock, Hans-Johann, and John Hyman, eds. *A Companion to Wittgenstein*. Chichester: John Wiley & Sons, 2017.

Godfrey-Smith, Peter. *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press, 2013.

Gontier, Nathalie, Jean Paul Van Bendegem, and Diederik Aerts, eds. *Evolutionary Epistemology, Language and Culture: A Non-Adaptationist, Systems Theoretical Approach*. Dordrecht: Springer, 2006. https://link.springer.com/book/10.1007/1-4020-3395-8.

Gould, S. J., R. C. Lewontin, J. Maynard Smith, and Robin Holliday. ‘The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme’. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205, no. 1161 (21 September 1979): 581–98. https://doi.org/10.1098/rspb.1979.0086.

Hjortland, Ole Thomassen. ‘Anti-Exceptionalism about Logic’. *Philosophical Studies* 174, no. 3 (2017): 631–58. https://doi.org/10.1007/s11098-016-0701-8.

Horsten, Leon. ‘Philosophy of Mathematics’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2022. Metaphysics Research Lab, Stanford University, 2022. https://plato.stanford.edu/archives/spr2022/entries/philosophy-mathematics/.

Hull, David L. ‘Chapter 2: Interactors versus Vehicles’. In *The Role of Behavior in Evolution*, edited by Henry C. Plotkin, 19–50. Cambridge, Mass: The MIT Press, 1988. https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=48504&site=ehost-live.

James, William. ‘Great Men, Great Thoughts, and the Environment’. *Atlantic Monthly* XLVI, no. CCLXXVI (1880): 441–59.

Kay, Paul, and Richard S. Cook. ‘World Color Survey’. In *Encyclopedia of Color Science and Technology*, edited by Ronnier Luo, 1–8. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015. https://doi.org/10.1007/978-3-642-27851-8_113-10.

Kay, Paul, and Luisa Maffi. ‘Color Appearance and the Emergence and Evolution of Basic Color Lexicons’. *American Anthropologist* 101, no. 4 (1999): 743–60. https://doi.org/10.1525/aa.1999.101.4.743.

Kirby, Jeremy. ‘Chrysippus’. In *Internet Encyclopedia of Philosophy*. Accessed 28 February 2022. https://iep.utm.edu/chrysipp/.

Kneale, William Carvert, and Martha Kneale. *The Development of Logic*. 6th edition. Oxford: Clarendon Press, 1975.
Kraut, Richard. 'Plato'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2022. Metaphysics Research Lab, Stanford University, 2022. https://plato.stanford.edu/archives/spr2022/entries/plato/.
Laland, Kevin, Blake Matthews, and Marcus W. Feldman. 'An Introduction to Niche Construction Theory'. *Evolutionary Ecology* 30, no. 2 (April 2016): 191–202. https://doi.org/10.1007/s10682-016-9821-z.
Laland, Kevin N., and Michael J. O'Brien. 'Cultural Niche Construction: An Introduction'. *Biological Theory* 6, no. 3 (September 2011): 191–202. https://doi.org/10.1007/s13752-012-0026-6.
Laland, Kevin N., John Odling-Smee, William Hoppitt, and Tobias Uller. 'More on How and Why: Cause and Effect in Biology Revisited'. *Biology and Philosophy* 28, no. 5 (September 2013): 719–45. https://doi.org/10.1007/s10539-012-9335-1.
Laland, Kevin N., Tobias Uller, Marcus W. Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, and John Odling-Smee. 'The Extended Evolutionary Synthesis: Its Structure, Assumptions and Predictions'. *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1813 (2015). https://doi.org/10.1098/rspb.2015.1019.
Levins, Richard. 'Coexistence in a Variable Environment'. *The American Naturalist* 114, no. 6 (December 1979): 765–83. https://doi.org/10.1086/283527.
Levins, Richard, and Richard Lewontin, eds. 'Adaptation'. In *The Dialectical Biologist*, 65–84. Harvard University Press, 1985.
Lewens, Tim. 'Cultural Evolution'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/sum2020/entries/evolution-cultural/.
Lewontin, Richard Charles. 'Models, Mathematics and Metaphors'. *Synthese* 15, no. 2 (1963): 222–44.
———. 'Organism and Environment'. In *Learning, Development, and Culture: Essays in Evolutionary Epistemology*, edited by H. C. Plotkin, 151–70. New York: John Wiley & Sons, 1982.
Lorenz, Konrad. *Behind the Mirror: A Search for a Natural History of Human Knowledge.* London: Methuen, 1977.
Moyal-Sharrock, Danièle. 'Wittgenstein on Knowledge and Certainty'. In *A Companion to Wittgenstein*, edited by Hans-Johann Glock and John Hyman, 545–62. Chichester: John Wiley & Sons, 2017.
Nails, Debra. 'Socrates'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/spr2020/entries/socrates/.
Odenbaugh, Jay. 'Idealized, Inaccurate but Successful: A Pragmatic Approach to Evaluating Models in Theoretical Ecology'. *Biology & Philosophy* 20, no. 2–3 (March 2005): 231–55. https://doi.org/10.1007/s10539-004-0478-6.
Pedersen, Olaf. *The First Universities: Studium Generale and the Origins of University Education in Europe*. Cambridge: University Press, 1997.
Pigliucci, Massimo. 'Stoicism'. In *Internet Encyclopedia of Philosophy*. Accessed 28 February 2022. https://iep.utm.edu/stoicism/.
Popper, Karl Raimund. 'Evolutionary Epistemology'. In *Evolutionary Theory: Paths into the Future*, edited by Jeffrey W. Pollard. Chichester West Sussex: Wiley, 1984.
Priest, Graham. 'Revising Logic'. In *The Metaphysics of Logic*, edited by Penelope Rush, 211–23. Cambridge: Cambridge University Press, 2014. https://doi.org/10.1017/CBO9781139626279.016.
Rawson, Glenn. 'Plato's Meno'. In *Internet Encyclopedia of Philosophy*. Accessed 21 February 2022. https://iep.utm.edu/meno-2/.

Robinson, Howard. ‘Substance’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/fall2021/entries/substance/.

Rogers, Kara. ‘Horizontal Gene Transfer’. In *Encyclopedia Britannica*, 22 August 2019. https://www.britannica.com/science/horizontal-gene-transfer.

Rosenblueth, Arturo, and Norbert Wiener. ‘The Role of Models in Science’. *Philosophy of Science* 12, no. 4 (October 1945): 316–21. https://doi.org/10.1086/286874.

Schroeder-Heister, Peter. ‘Proof-Theoretic versus Model-Theoretic Consequence’. In *The Logica Yearbook 2007*, edited by Michal Peliš, 187–200. Prague: Filosofia, 2007.

Shapiro, Stewart. ‘Logical Consequence, Proof Theory, and Model Theory’. In *Oxford Handbook of Philosophy of Math and Logic*, 651–70. Cary, UNITED STATES: Oxford University Press, Incorporated, 2005. http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3052016.

Shapiro, Stewart, and Teresa Kouri Kissel. ‘Classical Logic’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/spr2021/entries/logic-classical/.

Shields, Christopher. ‘Aristotle’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/fall2020/entries/aristotle/.

Smith, Robin. ‘Aristotle’s Logic’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/fall2020/entries/aristotle-logic/.

Whiting, Daniel. ‘Language, Language-Games, and Forms of Life’. In *A Companion to Wittgenstein*, edited by Hans-Johann Glock and John Hyman, 420–32. Chichester: John Wiley & Sons, 2017.

Williamson, Timothy. ‘Abductive Philosophy’. *The Philosophical Forum* 47, no. 3–4 (2016): 263–80. https://doi.org/10.1111/phil.12122.

Wittgenstein, Ludwig. *On Certainty*. Edited by Denis Paul and G. E. M. Anscombe. Oxford: Basil Blackwell, 1969.

———. *Philosophical Investigations*. Edited by G. E. M. Anscombe, R. Rhees, and G. H. Von Wright. 3rd edition. Oxford: Basil Blackwell, 1968.

Wuketits, Franz M. ‘Evolutionary Epistemology: The Non-Adaptationist Approach’. In *Evolutionary Epistemology, Language and Culture: A Non-Adaptationist, Systems Theoretical Approach*, edited by Nathalie Gontier, Jean Paul Van Bendegem, and Diederik Aerts, 33–46. Dordrecht: Springer, 2006. https://doi.org/10.1007/1-4020-3395-8.