Cognitive epistemology

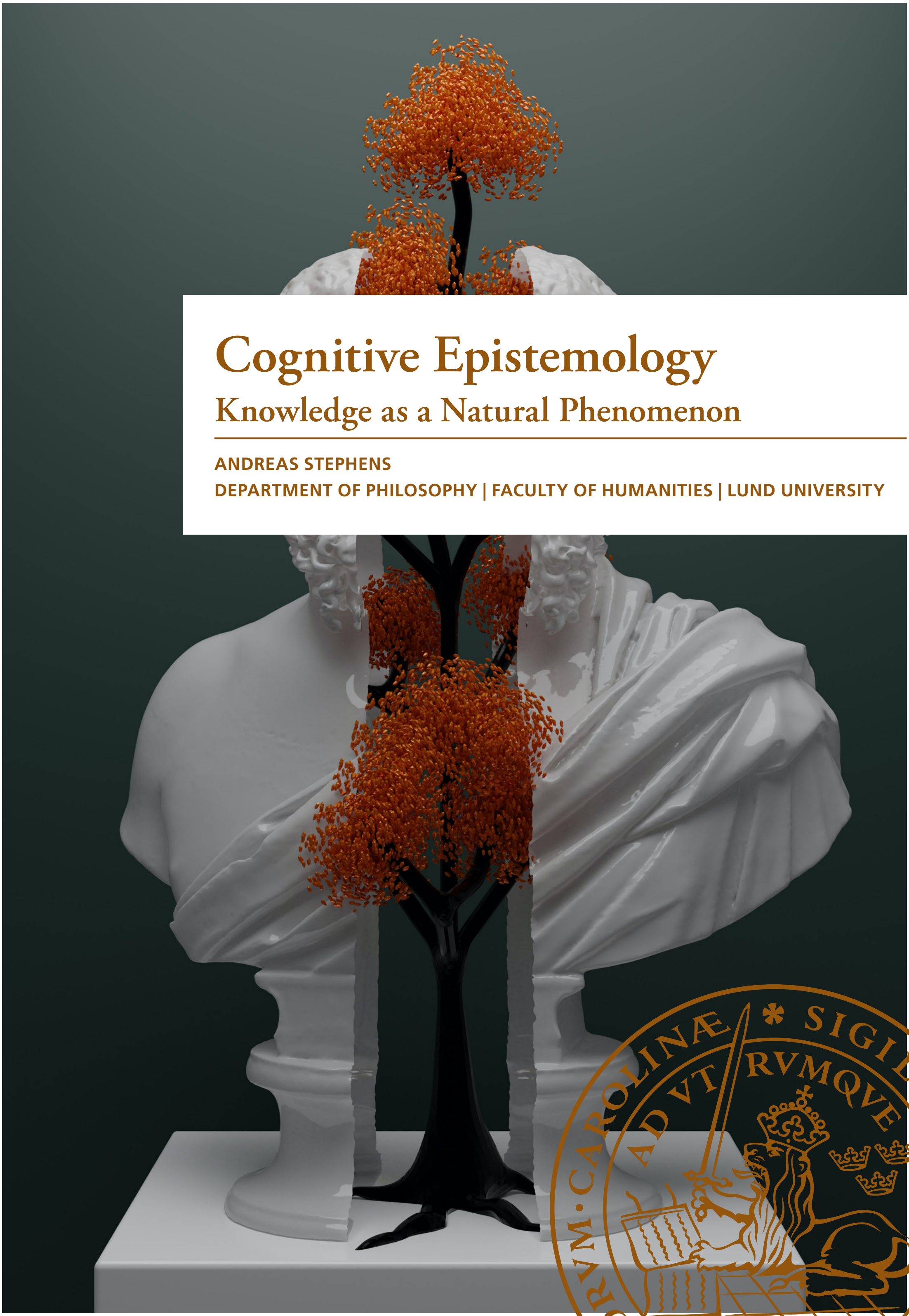
Knowledge as a natural phenomenon

Stephens, Andreas

2024

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Cognitive Epistemology

Knowledge as a Natural Phenomenon

ANDREAS STEPHENS
DEPARTMENT OF PHILOSOPHY | FACULTY OF HUMANITIES | LUND UNIVERSITY

Cognitive Epistemology

Cognitive Epistemology

Knowledge as a Natural Phenomenon

Andreas Stephens

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the
Department of Philosophy, Faculty of Humanities and Theology at
Lund University to be publicly defended on
Friday the 10th of May 2024 at 13:15 in LUX: B152.

Faculty opponent
Professor Klemens Kappel (University of Copenhagen)

Organization
LUND UNIVERSITY

Department of Philosophy
Box 192
SE-221 00 Lund, Sweden

Document name
DOCTORAL DISSERTATION

Date of issue
2024-05-10

Author(s)

Andreas Stephens

Sponsoring organization

Title and subtitle

Cognitive epistemology: Knowledge as a natural phenomenon

Abstract

This thesis investigates the question 'What is knowledge?' In intuition-based epistemology the question is often considered to concern how 'knowledge' is used linguistically or conceptually rather than what knowledge *is*. In addition, since intuitions are used as evidence despite empirical experiments indicating that people's intuitions vary a great deal and that little conclusive systematicity can be found, it is argued that approaches with such a focus cannot provide a solid foundation to answer the initial question. By instead looking at naturalistic approaches, a pluralistic cognitive epistemological approach which accepts ontological naturalism, methodological cooperative naturalism, and evolutionary epistemology can be identified. Given this approach – close to that of Hilary Kornblith – it is possible to look at how various relevant sciences see the natural phenomenon of knowledge. This provides a complement to Kornblith's sole focus on cognitive ethology. By also including the perspectives of cognitive psychology and evolutionary systems theory a new view of knowledge is made possible. The emerging picture indicates that the natural phenomenon of knowledge plausibly can be seen as consisting in dynamic internal survival-beneficial structures. For higher organisms, such structures importantly involve reflexive and reflective memory processes that (satisficingly) reliably produce (satisficingly) true beliefs.

Key words

Knowledge, Cognitive epistemology, Naturalistic epistemology, Cognitive psychology, Evolutionary systems theory

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language

English

ISSN and key title

ISBN

978-91-89874-02-2 (Print)

978-91-89874-03-9 (Digital)

Recipient´s notes

Number of pages

184

Price

Security classification

Distribution by (name and address)

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature ____________________ Date 2024-03-04

Cognitive Epistemology

Knowledge as a Natural Phenomenon

Andreas Stephens

LUND UNIVERSITY

Coverphoto by Mithushi Arya

Faculty of Humanities and Theology
Department of Philosophy
Lund University

ISBN: 978-91-89874-02-2 (Print)
ISBN: 978-91-89874-03-9 (Digital)

Printed in Sweden by Media-Tryck, Lund University
Lund 2024

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN

To Miles

Table of Contents

Acknowledgements

I would like to express my gratitude to my supervisors Martin L. Jönsson and Erik J. Olsson. Their encouragement, knowledgeable insights, and generous help made this thesis possible.

It has been a privilege to do my PhD at Lund University. Students, staff, and researchers have helped me in different ways, creating a great atmosphere. Thank you Henrik Andersson, Kristina Arnrup Thorsbro, Mark Bowker, Eric Brandstedt, Ingar Brink, Ingela Byström, Anna Cagnan Enhörning, Niklas Dahl, Dan Egonsson, Fredrik Eriksson, Cathrine V. Felix, Andrés Garcia, Emmanuel Genot, Mattias Gunnemyr, Frits Gåvertsson, Peter Gärdenfors, Tobias Hansson Wahlberg, Jan Hartman, Jana Holsanova, Hubert Hågemark, Birger Johansson, Petter Johansson, Marta Johansson Werkmäster, Asger Kirkeby-Hinrup, Pierre Klintefors, Gloria Mähringer, Jens Nirme, August Olsen, Kalle Palm, Carl-Johan Palmqvist, Johannes Persson, Tomas Persson, Philip Pärnamets, Toni Rønnow-Rasmussen, Signe Savén, Stefan Schubert, Eva Sjöstrand, Annah Smedberg-Eivers, Samantha Stedtler, Jakob Stenseke, Robin Stenwall, Alexander Tagesson, Betty Tärning, Gabriel Vogel, Ylva von Gerber, Annika Wallin, Jakob Werkmäster, Claudia Zeiträg, Frank Zenker, Anna Östberg, Fredrik Österblom, and all participants at the higher seminar in theoretical philosophy.

Outside of Lund I want to thank Carlo Cellucci, Julia Duetz, Mikkel Gerken, Hans-Johann Glock, Max Johannes Kippersund, Michael Morreau, and Martin Zach for collaborations and discussions.

I am particularly grateful for all my talks with Andrey Anikin, Per-Erik Camacho, Miguel Ganzo Mateo, Max Minden Ribeiro, Nina Poth, Maximilian Roszko, Arthur Schwaninger, Ylwa Sjölin Wirling, and Melina Tsapos.

Thank you Trond Arild Tjøstheim for all the inspiring conversations about philosophy, science, and art.

Lastly, a big thanks to my family and friends.

Abstract

This thesis investigates the question ‘What is knowledge?’ In intuition-based epistemology the question is often considered to concern how ‘knowledge’ is used linguistically or conceptually rather than what knowledge *is*. In addition, since intuitions are used as evidence despite empirical experiments indicating that people’s intuitions vary a great deal and that little conclusive systematicity can be found, it is argued that approaches with such a focus cannot provide a solid foundation to answer the initial question.

By instead looking at naturalistic approaches, a pluralistic cognitive epistemological approach which accepts ontological naturalism, methodological cooperative naturalism, and evolutionary epistemology can be identified. Given this approach – close to that of Hilary Kornblith – it is possible to look at how various relevant sciences see the natural phenomenon of knowledge. This provides a complement to Kornblith’s sole focus on cognitive ethology. By also including the perspectives of cognitive psychology and evolutionary systems theory a new view of knowledge is made possible. The emerging picture indicates that the natural phenomenon of knowledge plausibly can be seen as consisting in dynamic internal survival-beneficial structures. For higher organisms, such structures importantly involve reflexive and reflective memory processes that (satisficingly) reliably produce (satisficingly) true beliefs.

Populärvetenskaplig Sammanfattning

Denna avhandling söker ett svar på frågan ’Vad är kunskap?’ I traditionell epistemologi anses frågan ofta handla om hur begreppet ’kunskap’ används, snarare än att handla om vad kunskap faktiskt är. Då sådana tillvägagångssätt förlitar sig på intuitioner som evidens, trots att empiriska experiment visar att människors intuitioner varierar utan att någon grundläggande systematik har kunnat fastslås, kan de inte anses ge en solid grund för att svara på frågan om vad kunskap är.

Genom att istället undersöka olika former av naturalistisk epistemologi kan en pluralistisk kognitiv epistemologisk position presenteras. Kognitiv epistemologi anammar ontologisk naturalism, metodologisk kooperativ naturalism samt evolutionär epistemologi. Med denna utgångspunkt – vilken ligger nära Hilary Kornbliths – är det möjligt att undersöka hur olika relevanta vetenskaper ser på kunskap. Detta erbjuder ett komplement till Kornbliths undersökning, som endast inkluderar hur kunskap ses inom kognitiv etologi. Genom att även ta del av perspektiv från kognitiv psykologi och evolutionär systemteori kan en ny syn på kunskap nås. Det naturliga fenomenet kunskap kan då ses bestå i dynamiska inre överlevnadsfördelaktiga strukturer. För högre organismer involverar betydande sådana strukturer reflexiva och reflektiva minnesprocesser som på ett (tillräckligt) tillförlitligt sätt producerar (tillräckligt) sanna trosföreställningar.

List of Papers

This thesis includes the following papers.

Paper I

Stephens, A. (2016). **A pluralist account of knowledge as a natural kind.** *Philosophia*, *44*(3), 885-903. DOI: 10.1007/s11406-016-9738-3.

Paper II

Gärdenfors, P., and **Stephens, A.** (2017 [2018]). **Induction and knowledge-what.** *European Journal for Philosophy of Science*, *8*(3), 471-491.
DOI: 10.1007/s13194-017-0196-y.

Paper III

Stephens, A. (2019). **Three levels of naturalistic knowledge.**
In M. Kaipainen, F. Zenker, A. Hautamäki, and P. Gärdenfors (eds.), *Conceptual spaces: Elaborations and applications* (Synthese Library, Studies in Epistemology, Logic, Methodology, and Philosophy of Science, Vol. 405, pp. 57-73). Cham: Springer Nature Switzerland. ISBN: 978-3-030-12799-2, DOI: 10.1007/978-3-030-12800-5.

Paper IV

Stephens, A., and Tjøstheim, T. A. (2020 [2022]). **The cognitive philosophy of reflection.** *Erkenntnis*, *87*, 2219-2242. DOI: 10.1007/s10670-020-00299-0.

Paper V

Stephens, A., Tjøstheim, T. A., Roszko, M., and Olsson, E. J. (2021).
A dynamical perspective on the generality problem. *Acta Analytica*, *36*(3), 409-422. DOI: https://doi.org/10.1007/s12136-020-00458-6.

List of Published Papers Not Included in the Thesis

Stephens, A., (2023). Contextual shifts and gradable knowledge. *Logos & Episteme*, *XIV*(3), 323-337.
DOI: https://doi.org/10.5840/logos-episteme202314324.

Olsson, E. J., Tjøstheim, T. A., Stephens, A., Schwaninger, A., and Roszko, M. (2022 [2023]). The cognitive basis of the conditional probability solution to the value problem for reliabilism. *Acta Analytica*, *38*, 417-438.
DOI: 10.1007/s12136-022-00533-0.

Stephens, A. (2021). Consistency and shifts in Gettier cases. *Logos & Episteme*, *XII*(3), 327-339. DOI: https://doi.org/10.5840/logos-episteme202112324.

Tjøstheim, T. A., and Stephens, A. (2021 [2022]). Intelligence as accurate prediction. *Review of Philosophy and Psychology*, *13*, 475-499.
DOI: 10.1007/s13164-021-00538-5.

Stephens, A., and Felix, C. V. (2020). A cognitive perspective on knowledge how: Why intellectualism is neuro-psychologically implausible. *Philosophies*, *5*(3), 21. Reprinted in M. J. Schroeder, and G. Dodig-Crnkovic (eds.), *Contemporary natural philosophy and Philosophies—Part 2* (Philosophies Special Issue, pp. 33-46). Basel: MDPI. ISBN: 978-3-03943-535-7,
DOI: 10.3390/philosophies5030021.

Felix, C. V., and Stephens, A. (2020). A naturalistic perspective on knowledge how: Grasping truths in a practical way. *Philosophies*, *5*(1), 5. Reprinted in M. J. Schroeder, and G. Dodig-Crnkovic (eds.), *Contemporary natural philosophy and Philosophies—Part 2* (Philosophies Special Issue, pp. 47-57). Basel: MDPI. ISBN: 978-3-03943-535-7,
DOI: 10.3390/philosophies5010005.

Tjøstheim, T. A., Stephens, A., Anikin, A., and Schwaninger, A. (2020). The cognitive philosophy of communication. *Philosophies*, *5*(4), 39. Included in G. Dodig-Crnkovic, and M. J. Schroeder (eds.), *Contemporary natural philosophy and Philosophies—Part 3* (Philosophies Special Issue). Basel: MDPI.
DOI: 10.3390/philosophies5040039.

Cognitive Epistemology

Chapter 1

What is Knowledge?

WHAT IS KNOWLEDGE? As a starting point, let us look at this question from an everyday perspective. Seemingly, knowledge is important. Without it we risk losing money by placing bad bets, looking foolish by answering inappropriately, or failing to accomplish our tasks by messing up procedures. So even though it might be hard to pin down exactly what knowledge is or what it involves, we want it – indeed need it – in order to function in the world and in society. Taking this one step further, it might even be said that it is almost always a good thing to have a lot of knowledge, since the more knowledge we have, the better we fare, all other things being equal.

Moreover, we do seem to know a lot. On the one hand we know a lot of facts such as our telephone number and the names of capitals in various countries. On the other hand, we also know how to walk, swim, and ride a bicycle. Furthermore, we know what different things are, and what they are for. For example, we can tell a blue whale from a dolphin, we know what chopsticks and batons respectively are used for. We also know when certain historically or personally important events took place, or when they will take place – either with specific timing or more generally that they happened in the past or will take place in the future.

Now this being said, if we try to understand exactly what it is that takes place when we *know* something, the question immediately becomes difficult to answer. Unsurprisingly, there are thus different opinions concerning how we should view knowledge. In fact, the topic has been debated within philosophy for over two millennia, but the debates have failed to reach any consensus. There are several conflicting positions and approaches, as well as a number of problems and paradoxes that are often perceived to be important to solve in order to understand what knowledge is. What is clear is that knowledge is important (but see, e.g., Papineau 2021a). What is less clear is its nature. Unfortunately, it is also unclear what methods we should use to find out what it is (for comprehensive overviews see, e.g., Ichikawa and Steup 2018; Comesaña and Klein 2019; Sorensen 2020).

So how have philosophers traditionally investigated what knowledge is? This question will get different answers depending on where in history we look, but a common methodology involves introspection and an attempt to find a definition of knowledge, or an understanding of our usage of the concept, that matches our intuitions (see, e.g., Bealer 1992; Goldman 2007; Williamson 2007; Pust 2019). Another influential approach instead argues that we must look outwards into the world itself – importantly including a third-person perspective on our cognitive faculties – using an empiricist scientific methodology (see, e.g., Kitcher 1992; Kornblith 2002; Margolis and Laurence 2021, sect. 5; but see, e.g., Goldman 2007; Williamson 2007; Cappelen 2012).

Focusing on issues of methodology, much of present-day philosophy can, arguably, be illuminated by exploring how it compares to these two influential outlooks. This is what we will proceed to do in the following two background chapters. In chapter 2, we will first discuss some characteristics of traditional intuition-based epistemology which is aligned with the first outlook. Importantly, this will not be an exhaustive discussion of all – or even most – intuition-based approaches. Rather, some aspects and issues will be highlighted to facilitate the thesis' goals. In chapter 3, we then consider naturalism which is aligned with the second scientifically based outlook. Here some different takes on philosophy's connection to science will be addressed. Thereafter, in chapter 4, a specific naturalistic position – close to that of Hilary Kornblith – which we will call *cognitive epistemology* will be presented and endorsed. The delineation, and the following application, of this position amounts to the thesis' main goals. Specifically, we will assume that there is a natural phenomenon of knowledge, and that the sciences are our best source of input concerning what it is. Chapter 5 then explores how a pluralistic approach can provide input – albeit tentative and fallible – concerning the question of what knowledge is, as well as how this input might interact with Kornblith's knowledge-account. Finally, in chapter 6, we present some brief concluding remarks and the thesis' scientific publications (Papers I–V).

Chapter 2

Intuition-Based Epistemology

INTUITION-BASED EPISTEMOLOGY takes a central position in contemporary Western philosophical debate. In this chapter we will point out some aspects concerning a traditional tendency to understanding what knowledge is by working out how the concept of knowledge should be defined in light of intuitions concerning various thought experiments (Kitcher 1992; Ichikawa and Steup 2018; Margolis and Laurence 2021, sect. 5). Importantly, we will not offer a complete overview, or full presentation, of the many influential philosophical positions ranging from, for example, virtue accounts concerning intellectual properties of agents or communities (see, e.g., Zagzebski 1996) and knowledge first accounts where 'knowledge' is treated as a primitive (see, e.g., Williamson 2000), to Bayesian accounts focusing on degrees of belief or credences instead of knowledge (see, e.g., Lin 2023). Rather, to serve as a stepping-stone for the ensuing discussion, we will highlight some aspects and issues pertaining to those theories that do rely on intuitions:

> [... T]he project of analysing knowledge is to state conditions that are individually necessary and jointly sufficient for propositional knowledge, thoroughly answering the question, what does it take to know something? [...] It is not enough merely to pick out the actual extension of knowledge. Even if, in actual fact, all cases of *S* knowing that *p* are cases of *j*, and all cases of the latter are cases of the former, *j* might fail as an analysis of knowledge. For example, it might be that there are *possible* cases of knowledge without *j*, or vice versa. A proper analysis of knowledge should

> at least be a necessary truth. Consequently, hypothetical thought experiments provide appropriate test cases for various analyses [...]. (Ichikawa and Steup 2018, italics in original)

Even though there are missed nuances, this description, arguably, captures a common outlook. To exemplify, let us focus on factual knowledge(-that), setting other knowledge-forms to the side for now – which incidentally has been a common approach in modern philosophy. If we start with the influential justified-true-belief-definition (JTB) of knowledge as a heuristic, a subject *S* knows a proposition *p* iff she is justified in her true belief that *p* (e.g., 'The emergency telephone number in Sweden is 112.' or 'Paris is the capital of France.'). Sidestepping a number of details, from the intuition-based perspective, whether *S* knows that *p*, or not, will then depend on whether an evaluator finds it intuitively acceptable to attribute or ascribe justification and knowledge to the subject. Supporters of the particular definition will argue that evaluators should do so, while those opposing will strive to find problematic or paradoxical cases that make the definition seem unintuitive. Supporters will often counter, by trying to show that the objections are misconstrued and can be explained away. Or insist that the definition can be salvaged by some modification. Opponents will then likely disagree, seeking to strengthen their case against the intuitiveness of the definition. And so on. This highlights how it is important for philosophers in the intuition-based epistemological tradition, in addition to trying to find definitions for epistemologically relevant concepts such as 'justification' and 'knowledge,' to make sense of how people attribute or ascribe knowledge.

Taken together this means that proponents of this approach do not inquire into what knowledge *is*, but rather into how the concept is viewed, used, and defined.[1] Importantly, how 'knowledge' is used linguistically or conceptually is primarily investigated with a focus on what evaluators find intuitively acceptable (Itakura 2001).[2] Although this second inquiry might be interesting in its own right, it should be acknowledged as being separate from the first. Given this, it is

[1] At least it doesn't concern *directly* what knowledge is. It might do so given assumptions about a strong connection between attributions of knowledge and knowledge itself. But such assumptions are seldom spelled out.

[2] There are those who, like for example Cappelen (2012) and Deutsch (2015), deny that philosophers use intuitions as evidence at all. But it remains for them to convincingly defend this claim (for a critique of Cappelen and Deutsch see, e.g., Devitt 2015; for arguments in favor of the view that philosophers do in fact use intuitions as evidence see, e.g., Goldman 2007; Williamson 2007; Nagel 2012; Pust 2019).

important to settle matters concerning, for example, linguistic meaning, conceptual content, how attributions might work differently in different contexts, which perspectives are being used, and find a way to determine whose opinion is counted in or out.

But how should we think about the connection between intuitions concerning knowledge and knowledge itself? Kornblith (e.g., 2002, pp. 10–11) has argued that our intuitions concerning knowledge *typically* do pick out genuine and obvious instances of the phenomenon.[3] Using 'gold' as an example, someone's intuition might involve that all yellow rock-like lumps are gold. Typically, this intuition might very well help that person to pick out obvious instances of gold. However, human reflection is beset by a number of limitations and biases – a fact that Kornblith (2012) also has pointed out. So, the intuition that all yellow rock-like lumps are gold could involve mistakes. Either by identifying something as gold even though it is not (e.g., the yellow rock-like lump is pyrite – fool's gold) or by failing to identify an actual instance of gold (e.g., the gold nugget is for some reason not yellow). Importantly, it might most often not matter whether we choose to use 'gold' for Au (gold) or FeS_2 (fool's gold) in casual conversations. But in some contexts, such as for example the application of gold in electronics, it might be crucial to get the underlying mineral right or we might potentially get a lethal electrical shock. Likewise, in many contexts we are free to use 'knowledge' as we prefer. But in some contexts (given how knowledge is tied to the survival of organisms, which we will discuss in section 5.2 below) it might be crucial to get at the underlying phenomenon correctly.

Answers to the questions posed by intuition-based epistemologists have remained elusive. As a remedy to this elusiveness, many experimental philosophers nowadays seek to find answers about the folk-psychological conception of knowledge by taking steps towards a more statistically grounded methodology which probes test-subjects' intuitions (see, e.g., Alexander and Weinberg 2014; Knobe and Nichols 2017). They thus strive to use empirical means to find answers.[4] Moreover, they tend to focus on groups rather than on individuals. It is, however, important to be clear about just what such answers would amount to. The answers would still present how people intuitively understand 'knowledge,' possibly also relaying how they attribute, ascribe, and talk about knowledge. But,

[3] This, however, is an empirical claim – perhaps best suited for cognitive anthropology as Kornblith himself suggests.

[4] From an empiricist perspective, this does present a more robust way of reaching answers.

again, since our intuitions are fallible, there is no guarantee that they in fact get at the underlying phenomenon correctly.

So, what have the experimental philosophical results shown? Well, people's intuitions appear to differ a great deal, as well as be possible to influence (see, e.g., Nichols 2004; Talbot 2009; Machery 2017; Fischer and Sytsma 2021; but see, e.g., Cohnitz 2020; see also, e.g., Chalmers 2014; Margolis and Laurence 2021). Moreover, concerning several epistemological cases, some empirical findings indicate that there are systematic differences between groups (Knobe and Nichols 2008; Alexander 2012), while other findings deny that such systematic differences exist (see, e.g., Kim and Yuan 2015; Seyedsayamdost 2015a, 2015b; Turri 2016). These inconclusive results have led some theorists to argue that philosophers' intuitions ought to prevail since they (allegedly) are experts in these matters given their education and long history of grappling with the questions (Williamson 2011; Turri 2013; but see, e.g., Buckwalter 2016). Nevertheless, even if we were to accept this last claim, philosophers arguably have a notoriously hard time agreeing on any particular position, since even among philosophers intuitions differ. That is, regardless of whether one prefers to focus on the verdicts of individuals, groups, lay people, or philosophers: intuitions differ (Jackson 2011; Starmans and Friedman 2012; see also Deutsch 2015, p. 33).

Summing up, traditional intuition-based epistemology characteristically – although many exceptions exist – investigates knowledge with a focus on how the concept 'knowledge' is going to be defined, as well as how it is used linguistically and conceptually. Although intuitions might typically succeed at picking out obvious instances of knowledge, it is clear that intuitions about 'knowledge' differ in non-systematic ways. They might thus fail to relay relevant information if we are interested in the phenomenon itself. It might, of course, be interesting to understand how people report that they subjectively understand the concept of knowledge from a first-person point of view in specific circumstances, but it is important to acknowledge that this amounts to a separate question from finding out what knowledge (the phenomenon) really is.

Chapter 3

Naturalistic Epistemology

NATURALISTIC PHILOSOPHY is an influential approach to philosophy heeding an empiricist outlook. So, how is knowledge investigated in this tradition?

In short, the main methodological difference with this approach compared to that of the intuition-based approach is that it focuses on what a phenomenon *is* rather than on a particular *concept*, or on how people intuit, attribute, or talk about a phenomenon (see, e.g., Kornblith, 1995, 2002; Hawking and Mlodinow 2011; Cellucci 2014, 2017). So, the question 'What is knowledge?' is then to be answered by investigating the natural phenomenon – knowledge – occurring in the world.

For much of Western history, philosophy has been held in a particularly high regard, influencing how we see the world and our place in it. But since the development of modern science around the seventeenth century this position is no longer easily defended. Indeed, naturalists see the scientific method as the best way of reaching answers. This means that it is no longer obvious what philosophy has to offer, compared to what science provides. This situation has led to several naturalistic responses from philosophers.[5] From this point of view, philosophy

[5] It should be pointed out that, for example, Feyerabend (1975), Dupré (1993) and others highlight that it is difficult, or even impossible, to find *one* particular scientific method, a fact that is noteworthy and potentially problematic for any naturalistic approach. Other notable perspectives see it as philosophy's role to spawn new sciences (Cellucci 2014,

could be seen as being superfluous. The sciences arguably have a good grasp of what they are doing, and it is not obvious that philosophy can add anything of substance. However, most naturalistic philosophers argue that although philosophy is something distinct from science it can nevertheless remain useful. Philosophy should then, in some way or form, work in continuance with science.

In fact, many, if not most, philosophers nowadays would claim that they support some form of naturalism, if nothing else then at least the *ontological* naturalistic commitment that there is nothing otherworldly or 'supernatural' going on. A further commitment is made by *methodological* naturalism, which claims that philosophers need to heed a scientifically informed method. But what this ought to involve can pan out differently depending on how a number of details are interpreted (see, e.g., Kornblith 1993; Godfrey-Smith 2003; Feldman 2012; Papineau 2021b). After these initial remarks we will now discuss some influential naturalistic positions,[6] using Feldman's (2012) categorization as a starting point.

According to *replacement* naturalism, made famous by Quine (1969a), the inquiry concerning knowledge and justification is better taken over by science. This is so since it is science – not philosophy – that nowadays provides our best means of finding things out. Specifically, Quine argues, epistemology should leave the scene in favor of psychology:[7]

> The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology? [...] Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input—certain patterns of irradiation in assorted frequencies, for instance—and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for

2017) or to act as model-theoretician, creating new types of models for the sciences (Angere 2010).

[6] Primarily *methodological* naturalistic positions.

[7] We will not discuss the behaviorist psychological tradition that was influential in Quine's days, or whether/how his position might be translatable into a modern-day cognitive psychological setting.

> somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence. (Quine 1969a, pp. 75, 82–83)

This form of methodological naturalism is plausible since psychology indeed seems better suited to answer many of the questions traditionally posed by epistemologists. What knowledge is, is then best investigated by psychology. However, even though Quine's argument, which was primarily directed against a foundationalist form of philosophy, is accepted it is nevertheless possible to claim that the questions epistemologists pose lie outside the scope of psychology. This view has some merit in that psychologists might be uneasy in using the normatively laden outlook that has traditionally been used by philosophers, and so the replacement naturalistic position could be seen as just changing the subject (see, e.g., Kornblith 1995, p. 239).

According to *substantive* naturalism, the questions epistemologists pose need to be reformulated into a strictly scientific terminology (see, e.g., Churchland 1982). According to the substantive naturalist, the traditional method of relying on folk-psychological descriptions has dampened our prospect of progress. Terms such as, for example, 'justified,' 'warranted,' and 'knows that' need to give way to terms such as, for example, 'causes,' 'implies,' and 'believes that' (for a longer discussion see, e.g., Goldman 1979). By reformulating the issues at hand, we stand a chance of actually getting some answers. This methodology would indeed make epistemology fit better with science since all relevant terms would be reworked into scientifically acceptable ones. However, convincing translations have remained elusive. In comparison to replacement naturalism, someone who heeds substantive naturalism could agree with replacement naturalism that psychology would be a primary source of input for their investigation. The psychologist's vernacular is that which should be used. But they would not go as far as handing over their whole enterprise to the scientists. Arguably, the substantive naturalistic position is also susceptible to criticisms concerning changing the subject.

According to *cooperative* naturalism, epistemologists are free to ask any form of question they want (including normative ones) as long as they accept that scientific input overrides intuitive beliefs. That is, science is seen as providing our best understanding of the world and natural phenomena. So philosophical questions are potentially important in their own right, although they need to take relevant scientific findings and theorizing into account whenever such are to be had. This said, if such findings are lacking or no consensus is to be found, philosophers are encouraged to try to make headway on their own. Intuitions can thus fill an initial role, but they should be seen as being corrigible and we must realize that they are heavily influenced by our background beliefs. Thus, as soon as there is better input to be had, these intuitions must be abandoned for the new (better) scientifically grounded findings. Now, someone who heeds cooperative naturalism could – like the replacement naturalist – embrace psychology (as well as other sciences) and scientifically respectable problem formulations – like the substantive naturalist – but they would go further than the previous two naturalisms in their way to accommodate all questions of inquiry as being potentially fruitful while primarily relying on the results of science.

Yet another naturalistic position is offered by *evolutionary* epistemology[8] which highlights that natural selection has formed organisms (such as humans) into having particular cognitive faculties. This, it is claimed, must be taken into account to understand what a phenomenon such as knowledge amounts to (Sellars 1919; Campbell 1974; see also Plotkin 1993; Bradie and Harms 2020). We should acknowledge the fact that our understanding of the world is delimited by our body, importantly including our cognitive apparatus, which governs what we can and cannot process – it enables certain patterns to be experienced while ruling out others. This is perhaps easiest to illustrate by mentioning the vast differences that exist between various species' abilities. Notably, Cellucci (2017, p. 112) points out how our affordances and limitations in fact are crucial for our survival. If it were not the case that we were tuned to certain input, we would be utterly unfit to handle our environment with its particular offerings and threats that are paramount to our survival. We need to be able to handle 'mid-sized stuff' (relative to us humans) and are thus adapted to thrive in our niche of the world.

We have discussed how naturalism offers an alternative to intuition-based epistemology in its focus on what knowledge is. Knowledge is here understood as a natural phenomenon in the world that is best investigated using scientific means. There are, however, different methodologies concerning how philosophers should interact with science and its results. These range from claims that philosophy is better given up for the more advantageous pursuit that science manifests to claims that philosophy should be revamped in a more scientific manner, or that philosophy actually deals with issues that sometimes might lie outside the current interest of science which potentially makes philosophy remain a worthwhile venture, as long as relevant scientific input is heeded.

[8] We are here interested in EEM, focusing on the evolution of epistemological mechanisms working as a complementary source to philosophy, rather than in EET, focusing on the epistemological evolution of theories.

Chapter 4

Cognitive Epistemology

SO WHICH KIND of naturalistic epistemological position is the preferable one? In this section one possibly fruitful position will be described, that we will call 'cognitive epistemology.' The position is heavily influenced by Kornblith's discussions of epistemological methodology (see, e.g., Kornblith 1993, 1995, 1999, 2002, 2012, 2019, 2021). However, it involves some reinterpretations of Kornblith's position in the direction of pluralism. Cognitive epistemology endorses a methodological cooperative naturalistic stance, due to that stance's more inclusive position in that it does not disallow any inquiry out of hand. Rather, all questions are accepted as potentially interesting and important. Philosophers can then ask whichever questions they want but need to look to our best available theorizing for answers, i.e., science. In addition, an evolutionary epistemological perspective is also taken into account. This since cognitive epistemology, like Kornblith, emphasizes that the natural phenomenon of knowledge is tied to organisms with evolutionary histories.

Here is how Kornblith describes his view:

> On my view, knowledge is a natural phenomenon, and it is this natural phenomenon that is the subject matter of epistemology—not the concept of knowledge, but knowledge itself. Analyzing our concept of knowledge, to the extent that we can make sense of such a project, is no more useful than analyzing the ordinary concept of, say, aluminum. The ordinary

> concept of aluminum is of little interest for two reasons. First, most people are largely ignorant of what makes aluminum the kind of stuff it is, and so their concept of aluminum will tell us little about the stuff itself. Second, most people have many misconceptions about aluminum, and so their concepts of aluminum will reflect this misinformation as well. There are interesting anthropological questions about the ordinary concept of aluminum, but precisely because this concept is as much a reflection of ignorance and misinformation as it is a reflection of anything about aluminum, those who have an interest in aluminum are ill-advised to study our concept of it.
>
> Now the same may be said, I believe, of knowledge. Epistemologists ought to be interested in the study of knowledge itself. If we substitute [it for] a study of the ordinary concept of knowledge, we are getting at knowledge only indirectly; knowledge is thereby filtered through a good deal of ignorance about the phenomenon, as well as a good deal of misinformation. Better to examine the phenomenon of human knowledge in its natural setting and leave an examination of ordinary concepts to cognitive anthropology. The same may of course be said about justification and related epistemological notions. (Kornblith 1995, pp. 243–244)

Before we proceed to discuss the specifics of cognitive epistemology and Kornblith's position we need to address an important critique from intuition-driven theorists. In a commentary on Kornblith (2002), Goldman (2005; see also Olsson 2021, sect. 3) problematizes Kornblith's perspective by questioning how we can succeed at picking out the right natural kind and phenomenon out of all possible options without performing conceptual analysis first. Kornblith (2005, p. 430) discusses this point, arguing that it can be considered unproblematic (given his naturalistic perspective):

> Imagine an early chemist interested in the nature of acids. The term 'acid' was widely used before there was any real understanding of what it is that makes something an acid. So this chemist has vinegar (which is a dilute solution of acetic acid), hydrochloric acid, *aqua regia* (a mixture of hydrochloric acid and sulphuric acid) available in his laboratory, and he is trying to determine what, if anything, these various substances have in common. He believes they are all members of a single natural kind, and he is interested in determining what it is that makes them members of that kind. He has some views about what these substances have in common—many of which are mistaken—but instead of analyzing his concept of acid,

> he turns to the workbench and tries to figure out what these substances actually have in common. No one doubts the coherence of this project.
>
> Now imagine that another investigator hears about this project and announces that he wishes to help out. He too is going to find out what all acids have in common, and he has a number of samples of would-be acids which will form the basis of his investigation. Now suppose that the samples which this investigator is examining include shoes, ships, sealing wax and his pet dog. Clearly something has gone wrong. This second investigator is not engaged in the same project as the first, and it will be immediately obvious to anyone looking on that this is so. The same is true if this investigator has samples which are members of a single natural kind, but one nowhere in the vicinity of an acid: say, a dog, a cat, a cow and a sheep. How are we to explain the mistake that this investigator is making? (Kornblith 2005, pp. 429–430, italics in original, footnote removed)

Goldman argues that some form of semantico-conceptual analysis needs to be done here in order to separate the first project from the second. Kornblith agrees that there is something amiss in the second project pertaining to semantic competence:

> As I see it, individual investigators here must have a certain recognitional capacity in order even to begin: they must be able to recognize at least some samples of the stuff they wish to examine. I don't think that the proper way to understand this is by viewing this recognitional capacity as peculiarly semantic or conceptual, but, as I see it, this is not where the important issue is between Goldman and me. Suppose we say that this is a semantic or conceptual ability. The real issue is just how substantial the conceptual investigation must be. (Kornblith 2005, p. 430, fn. 4)

But, importantly, Kornblith only sees this as involving a rudimentary recognitional capacity and language competence, at most involving a trivial form of conceptual analysis:

> [... B]ut notice that the amount of conceptual analysis needed to rule out the bizarre or misguided investigator is utterly trivial. What is needed is not a detailed and fine-grained investigation of the concept of an acid; one certainly wouldn't want to devote two thousand years to arguing about the precise contours of the concept before ruling out these mistakes and getting on with the real work of studying acids. No such detailed investigation is necessary. (Kornblith 2005, p. 430)

This means that Kornblith (2005, pp. 430–431) considers that we should not engage in detailed analysis of irrelevant imaginary counterexamples and instead focus on the natural phenomenon aided by our rudimentary recognitional capacity (for critical discussions of Kornblith's view see, e.g., Goldman 2005; Kusch 2005; Talbott 2005; Talbot 2009; Olsson 2021; for some of Kornblith's replies see, e.g., Kornblith 2005).

Returning to the issue of the particular form of naturalistic epistemology that Kornblith subscribes to, it should be noted that he singles out cognitive ethology as *the* science that can give us an account of the natural phenomenon of knowledge (see also, e.g., Millikan 1984). He supports this view by arguing that '[k]nowledge, as it is portrayed in this literature, does causal and explanatory work.' (Kornblith 2002, pp. 28–29). So let us briefly look at Kornblith's interpretation of cognitive ethology (see Kornblith 2002 for his full knowledge-account):

> Cognitive ethologists are interested in animal knowledge precisely because it defines [...] a well-behaved category, a category that features prominently in causal explanations, and thus in successful inductive predictions. If we wish to explain why it is that members of a species have survived, we need to appeal to the causal role of the animals' knowledge of their environment in producing behavior which allows them to succeed in fulfilling their biological needs. Such explanations provide the basis for accurate inductive inference. The knowledge that members of a species embody is the locus of a homeostatic cluster of properties: true beliefs that are reliably produced, that are instrumental in the production of behavior successful in meeting biological needs and thereby implicated in the Darwinian explanation of the selective retention of traits. (Kornblith 2002, p. 62).

Kornblith reaches this reliabilist account of knowledge from how cognitive ethologists use intentional idioms, discussing a number of quotes from influential works in the field where knowledge is seen as a natural phenomenon and the term is used to describe said phenomenon. We will follow this methodology in chapter 5.

Now, a crucial issue on which cognitive epistemology diverges from Kornblith's approach is on the issue of pluralism. Kornblith's focus is solely on cognitive ethology, which he, rightfully, sees as providing relevant input to our understanding of the natural phenomenon of knowledge (see, e.g., Griffin 1976, 1978; Kingstone, Smilek and Eastwood 2008). This is indeed so, but as I will argue it is not the only relevant scientific account to be found.

Kornblith argues that it is often motivated to abstract away from the underlying details on lower levels of analysis (Kornblith 2002, pp. 39–42):

> There are commonalities among animals that can be captured at the level of talk of belief but cannot be captured in any lower-level vocabulary. A raven, for example, comes to believe that a hawk has been distracted, and thus attempts to steal its egg. Other ravens, similarly placed, behave in a similar way, and for much the same reason. *But there is no reason to think that the various ravens, each of which form a belief about some target hawk, have a common physical state in their brains.* There is, in particular, no more reason to think this about the ravens than there is to think this about human beings all of whom share a common belief. So we need to advert to some common property of the various individuals that abstracts from the details of the physical level of description. (Kornblith 2002, p. 41, italics added)

He continues his discussion by drawing attention to the difference between states involving informational content and those that do not:

> So when we look at a bit of animal behavior, one question we need to ask is whether its explanation requires talk of informational content, or whether some lower-level explanation, whether chemical or otherwise, will do. (Kornblith 2002, p. 41)

In a footnote Kornblith acknowledges that it is difficult to draw an explanatory line:

> Where to draw the line between those animals whose behavior can be explained in terms of sub-doxastic information-bearing states and those whose behavior can only be explained by a belief-desire psychology is a difficult empirical question. Drawing this line, however, is not necessary for the project of this book as long as it is clear, as I have argued, that the line does not place humans on one side and all other animals on the other. (Kornblith 2002, p. 42, fn. 16)

We can agree with Kornblith that abstractions that focus on informational states, beliefs, and desires indeed often are motivated. But we should nevertheless stress that such abstractions should be done with some caution. It is, according to cognitive epistemology, not appropriate to assume – like Kornblith does – that lower-level (or higher-level) explanations are irrelevant or insufficient. Kornblith's insistence that there is no reason to think that beliefs might have common lower-level properties – as the italicized portion of the above quote shows – is problematic since results from various (cognitive) sciences, arguably, do indicate that there are many commonalities between individual humans – as well as between species (Gazzaniga, Ivry and Mangun 2019). Technically, it might not be possible to find an absolute similarity between two *token*-states in different persons (or animals), but it is uncontroversial to claim that there are similarities between *type*-states. Cognitive neuroscientists commonly divide the brain into various modular functional categories that have particular roles and 'building blocks' (Panksepp 1998; Meunier, Lambiotte and Bullmore 2010). To exemplify, concerning hemispheric specialization it is commonly pointed out that '[h]umans, of course, have evolutionary ancestors, so we might expect to find examples of lateralized functions in other animals. Indeed, this is the case.' (Gazzaniga, Ivry and Mangun 2019, p. 159, see also pp. 160–163, 263). As another example to drive this point across, we can see how Gazzaniga, Ivry and Mangun (2019, p. 654), while discussing sentience, point out that '[...] subcortical brain areas arose early in the evolutionary process and are anatomically, neurochemically, and functionally homologous in all mammals that have been studied (Panksepp 2005).' Drawing a line between what should be included or abstracted away is thus more problematic than Kornblith concedes.

Moreover, we need to acknowledge, and take into consideration, many sciences – in theory possibly *all* – if we want a full understanding of any natural phenomenon given the world's complexity (Kusch 2005; Stephens 2016). That is, even though Kornblith's discussion of how knowledge is seen in cognitive ethology is highly interesting, cognitive epistemology encourages additional investigations. An examination of how other well-established sciences see the same natural phenomenon can arguably provide a richer picture. So, from the cognitive epistemological perspective, Kornblith's account should be understood as only providing a partial answer to the question of what knowledge is.

Numerous sciences inquire into cognition and knowledge. Many have their own perspective, working on their own level of analysis. So, it is not inconsequential to choose a particular science (or a particular selection of sciences) to focus on – natural phenomena can be modeled in different ways (see, e.g., Dupré 1993; Dupré and Nicholson 2018; see also, e.g., Marr 1982). But it is not, however, a *problem* from the perspective of cognitive epistemology. Rather, by looking at a natural phenomenon using different specific sciences, our understanding of the phenomenon arguably stands a chance to stepwise be improved.

Thus, cognitive epistemology encourages a multi-perspective triangulation of what knowledge is. By necessity each investigation will only involve a subset of all possible scientific perspectives. But regardless of where one's particular choice of focus is, other perspectives to start from should be acknowledged as being theoretically valid. Importantly, it is not the role of the philosopher – at least not the naturalistic philosopher – to decide which scientific perspectives and accounts should be accepted or not.

To clarify, according to cognitive epistemology it is fully reasonable of Kornblith to only focus on cognitive ethology (i.e., to choose one particular science to focus on) – but it is not reasonable of him to identify the account he finds as being the only (viable or relevant) one (i.e., to disallow input from other sciences):

> The conception of knowledge that we derived from cognitive ethology literature, a reliabilist conception of knowledge, gives us the only viable account of what knowledge is. (Kornblith 2002, p. 135)

Even though Kornblith writes this in order to make a specific argument that human knowledge is not different in kind from that of other animals', his book (Kornblith 2002) – taken as a whole – makes it plausible to consider him to stand by this claim more generally. However, more recently Kornblith (2021, p. 141) has softened his position concerning which sciences he believes are relevant for the investigation of knowledge to more broadly include '[...] the cognitive sciences, including psychology, neuroscience, linguistics, cognitive anthropology, cognitive ethology, and parts of sociology.' This is in line with the cognitive epistemological view.

Given our pluralistic commitment and given that we don't want to rely overly on exact language use or intuitions about what to call knowledge, we need to figure out under which circumstances we can assume that different theories are about the same natural phenomenon (we cannot naively take every scientific statement in terms of 'knows' or 'knowledge' to be about the natural phenomenon we are interested in or always be about the same thing). To facilitate the use of a variety of insights we need a 'not too rigid' method, since theoretical differences are to be expected, due to dissimilar focuses of perspective or level, as well as dissimilar abstractions or idealizations made in different models. As discussed in chapter 2, we will follow Kornblith (2005, p. 430) in assuming that our rudimentary recognitional capacity and language competence make us capable of comprehending convergence on objects of study in various scientific disciplines, as well as ruling out bizarre alternatives.

As a preliminary outset, we will assume that if our rudimentary language competence tells us that two scientific theories about phenomena *x* and *y* are about the same phenomenon (i.e., they both apply to a paradigmatic case of knowledge), and they characterize *x* and *y* in ways that are congruent, then it is reasonable to consider that they are about the same natural phenomenon. If both theories also use the same term to denote *x* and *y*, we will consider this as presenting additional corroborating support.

We can use Mitchell's (2002, 2003) idea of *integration* to clarify our point. Here scientific theories and models, that have been developed to capture phenomena in our complex world, are seen as being either non-competitive (compatible) or competitive (incompatible) where '[a]lmost all recent philosophers of science concerned with pluralism have concentrated exclusively on multiple, competing hypotheses, such as the wave and particle theories of light or Darwinian and Lamarckian theories of inheritance.' (Mitchell 2003, p. 208).

Importantly, at any given level, various competitive (incompatible) theories and models can be seen as presenting complementary explanations resulting from differences in choices concerning abstractions and idealizations – just as well as being genuinely competitive (incompatible). It is then primarily *cross-level* congruence that can help us differentiate between plausible and implausible theories and converge on an object of study (Mitchell 2002, 2003). This since the most plausible theories on each level will constrain what is plausible upwards or downwards hierarchical levels.

That is, if a certain theory (concerning, e.g., quanta or evolution) is compatible (non-competitive) with plausible theories on other levels of analysis (concerning, e.g., atoms or genetics), this would provide corroborating evidence of that theory's plausibility. Cross-level incompatibility (competitiveness) would instead indicate that something is amiss. For example, if the Darwinian theory of evolution (inheritance, with favorable/unfavorable variation, and natural selection governs evolution) is compatible (non-competitive) with our most plausible genetics, this would strengthen the plausibility of the Darwinian case. If, on the other hand, the Darwinian theory of evolution is incompatible (competitive) with our most plausible genetics this would weaken the plausibility of the Darwinian case (e.g., if genetics instead would be more readily compatible with Lamarckism; inheritance involves characteristics acquired during a parent organism's lifetime).

So, in situations where scientific inquiries have reached maturity, certain cross-level 'series' of theories might stand out as being more congruent – readily compatible – thus making them more plausible than others. Such convergence on objects of study is elucidated through '[i]ntegration [which] refers to a general class of scientific activity that does not directly involve either manipulation or observation, but focuses instead on hypothesizing, ordering, and cross-referencing connections between phenomena.' (Silva and Bickle 2009, p. 108, italics removed; see also Mitchell 2002, 2003).

Cognitive epistemology will thus follow Kornblith's methodology of looking to science for input, viewing how natural phenomena are characterized in particular sciences, as well as how they are denoted in said sciences. But where Kornblith identifies how cognitive ethology characterizes a certain natural phenomenon and denotes it 'knowledge,' cognitive epistemology encourages further exploration by involving more sciences, on other levels of analysis. Accordingly, we will present and discuss quotes from influential sources in other scientific fields than cognitive ethology. If congruent (compatible, non-competitive) characterizations are found between levels it would make it reasonable to consider that the different sciences are referencing the same natural phenomenon, as well as that their integrated accounts add plausibility to each other, and hence that they jointly give us a fuller picture of knowledge.

We have identified a specific epistemological position – cognitive epistemology – which accepts ontological naturalism, methodological cooperative naturalism, and evolutionary epistemology. Cognitive epistemology is influenced by Kornblith, although where Kornblith focuses solely on cognitive ethology, cognitive epistemology strives to get as much relevant scientific input as possible from multiple levels of analysis. We will in the next chapter therefore seek to identify a plausible cross-level series of theories of knowledge, which hopefully can inform us of complementary input concerning the natural phenomenon.

Chapter 5

Triangulating Knowledge

GETTING BACK TO our initial question 'What is knowledge?,' we will now choose two relevant well-established sciences, other than cognitive ethology (working on other levels of analysis), in order to complement Kornblith's account. Accordingly, in sections 5.1 and 5.2, we will try to illustrate how knowledge is characterized in cognitive psychology and evolutionary systems theory (EST) respectively, showing that '[k]nowledge, as it is portrayed in this literature, does causal and explanatory work.' (Kornblith 2002, pp. 28–29). Following Kornblith's lead, this will be done by presenting a discussion that revolves around representative quotes from influential sources in the two fields. In section 5.3, we will then argue that a plausible integration can be identified, which involves some possible ramifications for Kornblith's interpretation of knowledge.

5.1 From a Cognitive Psychological Point of View

We will in this section focus on cognitive psychology in order to offer some complementary input to Kornblith's cognitive ethological picture of knowledge. With the help of influential representative quotes from the field we aim to present an overarching story. So, how is knowledge characterized in cognitive psychology?

Cognitive psychology links knowledge to memory, where '[m]ore broadly, long-term memory is taken to constitute a person's knowledge of the world and this encompasses everything that they know.' (Quinlan and Dyson 2008, p. 356). Human memory is typically functionally understood as involving some form of ability to encode, store and retrieve information as mental representations. This form of computational metaphor allows us to understand and talk about memory systems without going deeper into neurobiology and neurochemistry. In reality, of course, memory depends on intricate neural structures and activations.

Canonical interpretations include Tulving's (see, e.g., 1985, 2002, 2005; see also Graf and Schacter 1985; Kandel et al. 2013) account of long-term memory (LTM) and Baddeley's (see, e.g., 2007) account of working memory (WM). Following the evolutionary development of memory capabilities in different species, Tulving partitions LTM into procedural, semantic, and episodic memory.

Procedural memory is evolutionarily prior, found in numerous species, governing perceptual and motor abilities, as well as aspects of cognitive skills:

> Procedural memory is proposed as the system containing knowledge of how to do things. This kind of knowledge guides both physical activities like cycling or swimming, and (partially) cognitive skills like playing chess or speaking in public. Usually, many trials are needed to acquire procedural knowledge, although one-trial learning does occur. These skills are hard to express verbally, if at all; the only way to show their presence is by means of performance. (Ten Berge and Van Hezewijk 1999, p. 607)

Using examples from the introduction in chapter 1, it is procedural memory that governs our knowledge about how to walk, swim, or ride a bicycle. Procedural memory is thought to be non-conscious. That is, it guides an agent's performance without her first-person conscious experience or grasp. We might know that we are able to walk, swim, or ride a bike. But we are not consciously aware of how we actually do it. Rather, this competence, that lets us retrieve and use relevant procedural memories, is something we train over multiple iterations of repeated action that enables automatic implicit activation. Put differently, after having identified key aspects of something we want to learn, we can practice it repeatedly which eventually lets us form adequate patterns of autonomous responses.

Semantic memory, also widely spread among animals, has evolved out of procedural memory. It governs sense-associations, generalized and conceptual understanding of the world such as knowing what things are and what they can be used for:

> How do we know what we know about the world? For instance, how do we know that a cup must be concave, or that a lemon is normally yellow

> and sour? Psychologists and cognitive neuroscientists use the term *semantic memory* to refer to this kind of world knowledge. […] Today, most psychologists use the term *semantic memory*—to refer to all kinds of general world knowledge, whether it be about words or concepts, facts or beliefs. What these types of world knowledge have in common is that they are made up of knowledge that is independent of specific experiences; instead, it is general information or knowledge that can be retrieved without reference to the circumstances in which it was originally acquired. (Yee, Chrysikou and Thompson-Schill 2014, p. 353, italics in original)

Semantic memory involves objective knowledge, in the sense that it is open to all (not just to a certain person). If we again go back to the examples found in the introduction, it is this memory type that enables us to tell a blue whale from a dolphin, as well as know what chopsticks and batons respectively are used for. But semantic memory also governs factual propositional knowledge such as the knowledge that 'The emergency telephone number in Sweden is 112.' or 'Paris is the capital of France.' which we saw as typical illustrations in connection to our discussion of the influential JTB-account of knowledge in chapter 2 above. Another central aspect of semantic memory is categorization, which involves hierarchical structures where we classify our knowledge in superordinate and subordinate categories.

The last of the three nested memories – episodic memory – is only found to a high degree in humans (although many other species are considered to have it to lesser degrees) (see, e.g., Dere et al. 2006; Templer and Hampton 2013).[9] It governs a sense of agency and remembrance such as knowing that something specific happened yesterday and being able to actively recall the fact. This means that it is contextual, involving phenomenological experiential features while being focused on the person in question:

> Semantic memory is conceptually based knowledge about the world, including knowledge of people, places, the meaning of objects and words. It is culturally shared knowledge. By contrast, episodic memory refers to memory of specific events in one's own life. The memories are specific in time and place. For example, knowing that Paris is the capital of France is

[9] We will not venture into the debate of whether it is more prudent to reframe episodic-like memory in animals (even though there are underlying neural similarities) as something completely different from human episodic memory.

> semantic memory, but remembering a visit to Paris or remembering being taught this fact is episodic memory. (Ward 2010, p. 186)

As can be seen in the above quote, episodic memory has autobiographical qualities. When we mentally 'time travel,' that is when we recall a particular situation that we were a part of, we use episodic memory. But episodic memory not only allows us to remember the past, it also, and arguably more importantly, lets us predict the future. This means that episodic memory allows us to become a little wiser by imagining possible scenarios without actually having to perform them in real life. This is advantageous. For example, if I have touched the fire once, I can recall the episode thereby avoiding making the same mistake again. Moreover, others who saw my accident can use their episodic memory of the event to predict likely scenarios concerning their own actions thereby avoiding making the same mistake as I did. Such predictions might not be perfect, but they, generally, need to be good enough to ensure survival. Episodic memory thus allows us to see ourselves as an active agent in a 'story,' which enables us to plan our actions in accordance with longer time-preferences. For example, even though we have found something to eat while being hungry, we can see a likely scenario where it is better to save some for later rather than to eat everything now. In taking such actions we enhance our overall survival chances. To once more use the examples from the introduction, episodic memory governs knowledge concerning when certain historically or personally important events took place, or when they will take place – either with specific timing or more generally that they happened in the past or will take place in the future.

LTM works in close contact – via episodic LTM – with WM, consisting of a central executive, a phonological loop, a visuospatial sketchpad, and an episodic buffer (Baddeley 2007). The central executive is thought to govern attentional and executive control functions, also coordinating the processes of the sub-systems. This overarching central system thus helps us maintain order. This can be thought of as a form of supervisor system, but can also be criticized since it seemingly works as a funnel for all phenomena that we are not able to explain yet (Quinlan and Dyson 2008, p. 379). The phonological loop sub-system governs verbal functions. This includes the comprehension as well as the production of speech. The visuospatial sketchpad governs visual management. That is, this sub-system maintains visual and spatial information over shorter timeframes. The episodic buffer governs cross-domain information linking. This sub-system thus helps store temporary information while also integrating different forms of input. In short, WM handles current short-term cognitive processing, enabling complex cognition and action.

The above discussion can be interpreted as indicating a tripartite knowledge-account consisting in knowledge-how (procedural memory), knowledge-what (semantic memory), and knowledge-that (episodic memory and WM) (Gärdenfors and Stephens 2017; Stephens 2019; Stephens and Tjøstheim 2020).

There are, however, alternative views. For example, one influential alternative categorizes memory into a non-declarative (procedural) and a declarative form:

> We propose that perceptual-motor and pattern-analyzing skills belong to a class of operations governed by rules or procedures; these operations have information-processing and memory characteristics different from those operations that depend on specific, declarative, data-based material. Although the distinction we have drawn between these classes of information may not permit all tasks to be sharply dichotomized, it should prove useful in predicting what is affected or spared in amnesia. This distinction between procedural or rule-based information and declarative or data-based information, which is reminiscent of the classical distinction between "knowing how" and "knowing that," has been the subject of considerable discussion in the literature of cognition and artificial intelligence [...]. The experimental findings described here provide evidence that such a distinction is honored by the nervous system. (Cohen and Squire 1980, p. 209)

This bipartition of memory/knowledge into a non-declarative (procedural) and a declarative form matches a united interdisciplinary consensus being formulated as, for example, a *standard model of the mind* (SMM) or a *common model of cognition* (CMC) (see, e.g., Laird, Lebiere and Rosenbloom 2017; Steine-Hanson, Koh and Stocco 2018; Stocco et al. 2018; Stocco et al. 2021). Even so, exactly how semantic and episodic memory are to be categorized remains an open question:

> In addition to facts, declarative memory can also be a repository of the system's direct experiences, in the form of episodic knowledge. There is not yet a consensus concerning whether there is a single uniform declarative memory or whether there are two memories, one semantic and the other episodic. The distinction between those terms roughly maps to semantically abstract facts versus contextualized experiential knowledge, respectively, but its precise meaning is the subject of current debate. (Laird, Lebiere and Rosenbloom 2017, p. 22)

Our chosen perspective presents an initial characterization of what knowledge is and how it works. Knowledge, from the perspective of cognitive psychology, is a natural phenomenon, i.e., our memory systems, which involve the above-described subsystems/components and their workings. Two plausible influential interpretations that explicitly link memory to knowledge stand out. A tripartite knowledge-account consisting in knowledge-how (procedural memory), knowledge-what (semantic memory), and knowledge-that (episodic memory and WM), as well as a bipartite knowledge-account consisting in knowledge-how (non-conscious and automatic procedural non-declarative memory) and knowledge-that (conscious declarative memory).

5.2 From an Evolutionary Systems Theoretical Point of View

To get additional input concerning what knowledge is, let us now look at how knowledge is characterized in EST, having close ties to, for example, cybernetics and systems science (see, e.g., von Bertalanffy 1968; Laszlo 1972a, 1972b; Badcock 2012; Ramstead, Badcock and Friston 2018a, 2018b; Badcock et al. 2019). Once more, this will be accomplished with the help of representative quotes from the field in focus. So, how is knowledge characterized in EST?

Ramstead, Badcock and Friston (2018a, p. 2) highlight how 'EST is an interdisciplinary field that [...] explains dynamic, evolving systems in terms of the reciprocal relationship between general selection and self-organisation.' This is accomplished by synthesizing different perspectives, striving to merge both ultimate and proximate perspectives (Tinbergen 1963):

> [...] around four specific, interrelated levels of analysis: functional explanations for evolved, species-typical characteristics; explanations for between-groups differences arising from phylogenetic mechanisms; explanations for individual differences resulting from ontogenetic processes; and mechanistic explanations for real-time phenomena, respectively. (Badcock 2012, p. 10)

EST is thus a metatheory that ties knowledge closely to biology, since it is only when we grasp how living organisms have evolved and come to understand their world that we will get a fuller comprehension of what knowledge truly encompasses. This approach strives to complement reductionist atomistic approaches through its holistic dynamical focus. This means that it seeks to include many different sciences and fields in its investigations, hoping that the bridging can offer new interdisciplinary insights. An issue that becomes apparent, from this perspective, is the risk of trying to understand a natural phenomenon that is processual and dynamic, by way of static states. This will undoubtedly result in problematic conceptualizations since no matter where 'the lines' are drawn; it will be a mere timeslice of a larger complex process involved in feedback loops.

According to EST all life-forms depend on knowledge to cope with the world:

> [... A] fly, a dog or a human being has only limited knowledge of the world, but [...] this knowledge has some validity because otherwise the fly, the dog, the human would not have been able to survive for long. (Hofkirchner 2005, sect. 1.3)[10]

Knowledge, then, is not something that happens in language or in a vacuum. It is a natural phenomenon, in actual organisms (Stephens et al. 2021).

Organisms can be said to constitute a form of 'whole.' They are not just the sum of their parts. Instead, the relationship between the various parts is of utmost importance. Individual parts can thus be replaced (and are indeed constantly replaced) while it is still reasonable to talk about the same organism. Laszlo (1972b, p. 70) describes this as that '[...] the organism is a "constitutive" (non-summative) totality of interdependent components.' In a sense, organisms are orderly processes that take place over their lifetimes. This since organisms struggle against the dissipative world. That is, they manage to keep themselves orderly and stable in a world with environmental entropic pressures (Schrödinger 1944). In other words, organisms must remain in states with low entropy (disorder) (Badcock, Friston and Ramstead 2019). If they find themselves outside such states, they must endeavor to get back quickly or else risk death. By ingesting energy and removing waste, organisms through their metabolism can remain in a negentropic (orderly) state that allows them to live on:

> Generally, knowledge is essential to the life of all organisms. For, in order to be and stay alive, all organisms must [...] explore the ecological possibilities available to them, and to this end they must have knowledge of the environment. (Cellucci 2022, p. 420)

The overall entropy in the environment is, however, increased by their behavior. In short, organisms need to be able to keep themselves – and their parts – alive and in check, as well as have a repertoire of behaviors at their disposal to tackle

[10] Hofkirchner gives Davidson (1983) as a reference.

unforeseen environmental events. A failure to do so can lead to dispersion of the organism, due to internal factors (organs might fail) or external factors (the organism might be consumed) (Kruglanski, Jasko and Friston 2020). Crucially, feedback from the world, as well as adjustments due to the feedback, gives the organism a possibility to remain in homeostasis (a steady-state equilibrium).[11] In relation to homeostasis it can be said that the organism needs to reduce the variety of the states it finds itself in to function properly. When it finds itself outside its equilibrium it quickly needs to take action to get back. So, the organism exploits '[...] the energy and structure of its environment for its growth and stability' (Sayre 1976, p. 116). To be specific, it is the flexibility of said exchange that is of primary importance in enabling organisms to thrive (Wiener 1948; Ashby 1956; Sayre 1976).

Organisms have accordingly adapted so that it is reasonable to expect to find them in low entropy states. Importantly, there is generally a high probability that they will remain there, '[...] self-organising systems that can avoid surprising phase-transitions have been favoured by natural selection over those that could not.' (Badcock, Friston and Ramstead 2019, p. 108). Organisms have thus certain inherited expectations that are hardwired in accordance with their species' evolution – the rest about their environment needs to be learned:

> [... T]he structure of the brain recapitulates the structure of the world in which it is embedded: environmental causes that are statistically independent are encoded in functionally and anatomically segregated neuronal structures. Similarly, the hierarchical organisation of the brain mirrors the hierarchically nested structure of causal regularities in the environment. This hierarchical nesting marries the hierarchy of temporal scales at which representations evolve with the hierarchy of temporal scales at which biological phenomena unfold – the lower, more peripheral layers of the neural hierarchy encode rapid environmental fluctuations associated

[11] There is an ongoing discussion about whether it is more precise to view the process of reaching such a state as involving 'homeostatic mechanisms' or 'allostasis.' Those preferring the allostatic formulation point out that organisms actually never are in a steady-state, a 'perfect' equilibrium, or in homeostasis. Rather they can be thought of as being involved in ongoing non-equilibrium allostatic processes of predicting and regulating their needs. Those preferring the homeostatic formulation tend to consider these points to be compatible with, and already included in, their outlook (see, e.g., Cannon 1929; Sterling and Eyer 1988; Berridge 2004; Carpenter 2004; Day 2005; McEwen and Wingfield 2010; Sterling 2012; Corcoran and Hohwy 2018). In what follows we will use the homeostatic formulation.

> with sensorimotor processing and stochastic effects; its higher, more central layers encode increasingly slower regularities related to contextual changes. (Badcock, Friston and Ramstead 2019, p. 108)

Sidestepping *internal* homeostatic processes, such as for example those involving body chemistry, the majority of organisms rely solely on automatic or instinctive behavior, although some have developed more elaborate means to survive and thrive in their more complex environments. That is, for organisms that live in environments where they can stay alive long enough to foster offspring, without having higher thought or intelligence, instinct is indeed enough. But for those organisms that live in environments that are more dynamic, prone to rapid change while involving complex obstacles, intelligence is needed. So, various degrees of intelligence have evolved in a number of species, whereby they embody a fit between themselves and their complex environment. Reformulating this point slightly, the organisms that live short lives in stable environments can afford mostly hardwired responses, whereas organisms that live longer lives, in more complex environments, need to be more flexible and learn. In this sense the world makes sure that our understanding is 'good enough' since organisms that fail to track the world will likely die before reproducing, while an unnecessarily rich understanding will likely lead to an unsustainable expenditure of energy (Simon 1955, 1956; Cellucci 2017; Artinger, Gigerenzer and Jacobs 2022).[12]

Higher organisms need to acquire proper mental states about relevant states of affairs in order to survive and function in the dynamic world they inhabit. This higher flexibility is enabled through intelligence and learning, which helps the organism cope in their environment. In doing so, organisms become able to assert some control (Wiener 1948; Ashby 1956; Friston, Kilner and Harrison 2006; Friston 2009, 2010) using their cognitive abilities, even though they are beset by limitations and biases (e.g., process speed is not infinite, only a couple of things can be held in mind at the same time etc.). Such abilities include many embodied features but also internal, as well as external, regulation.

Returning specifically to knowledge, Mobus and Kalton (2015; see also, e.g., Turchin 1993; Plotkin 1993; Mobus 2022) offer an illuminating interpretation that ties knowledge to *structure*:

[12] This same point is famously expressed as follows by Quine (1969b, p. 126): 'Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind.'

> We might consider knowledge as the cumulative expectations with which a system moves into the future. In this sense we say that the system *knows* what it needs to know in order to exist comfortably in the flows of its current environment. Knowledge, then, is that internal structure in a system that matches its capacity to dissipate flows in a steady-state equilibrium. We use the steady-state example here to make clear that knowledge is the fit between structure-grounded expectation and the actual situation as it unfolds. (Mobus and Kalton 2015, p. 297, italics in original)

This is a general description thought to be applicable to all systems. We should here acknowledge that this systems-perspective, arguably, diverges from the common philosophical anthropocentric focus on knowledge.[13] However, even though humans might be unique in many ways, we do share an evolutionary history with all life on Earth. Viewing us as dethatched from nature or ignoring

[13] Mobus and Kalton are aware that their perspective might be thought of as being idiosyncratic: 'This is a highly abstract, narrowly functional approach to knowledge. Most people think of knowledge as something that one possesses about the nature of the world that helps them navigate successfully in that world. We are looking with special focus on the navigation component of expectation, its functionality in enabling a system to move along handling the future with systemic adequacy. Nonconscious metabolic components have this kind of knowledge, and failures of knowledge as well, as in the cases where overactive immune system causes life-threatening allergies. And even our conscious forms of knowledge are grounded, like metabolisms, in physical structures and flows, for they are actually embodied in the intricacies of the connections of neurons in your brain. When you learn something new, by receiving information that is about something which affects you, the neurons in your brain literally undergo some rewiring or at least a strengthening of some existing connections. When you receive information in the strict sense that we have defined it above, the actuator biochemistry in your brain cells goes to work to generate new axonal connections where needed to support the representation of what you now know.' (Mobus and Kalton 2015, pp. 297-298). Saying that non-conscious metabolic components have knowledge might not be in line with how 'knowledge' is most commonly used. However, as we have argued above, scientific evidence about natural phenomena takes precedence over intuitions about how to use 'knowledge.' It can also be mentioned that this widening of knowledge is compatible with a rudimentary recognitional capacity concerning 'knowledge' since even though the EST-approach to knowledge makes knowledge more widespread than intuition might have it, it is still the case that many prototypical cases of knowledge are classified as knowledge.

the evolutionary aspect of the natural phenomenon of knowledge thus risks leading to an incomplete understanding of what knowledge is. Mobus and Kalton's broader notion of knowledge instead has the potential fruitfulness of providing a framework that can naturally explain and place human knowledge in a larger biological setting. This perspective lets us 'zoom out' and see what we have in common with other species. It also lets us 'zoom in' and see what makes us unique. Furthermore, it is pluralistic in the sense that it lets us focus on a particular aspect of knowledge while still being able to acknowledge other aspects as being important.

Let us now try to unpack Mobus and Kalton's description by choosing humans as our system of interest (to see what human knowledge consists in). The system's boundary can then be seen as consisting of our skin. The internal structure is thus the structure of our body (importantly including our cognitive faculties), shaped both by our genes and by what we experience in our lives. This means that we (as an evolved species) are affected over phylogenetic timescales, as well as over ontogenetic timescales (as individuals). The internal structures that amount to knowledge are the ones that help us interact with the world in a prosperous way, thus helping us to stay alive (personally) and procreate (stay alive as a species). That is, the internal structures that enable us to anticipate and tackle whatever the world throws at us (dissipate flows), letting us remain 'stable' and live on (in a steady-state equilibrium). This 'stability,' of course, involves an ongoing dynamic process of eating, drinking, breathing, and defecating, as well as solving various problems such as finding food and mates, avoiding predators and bad weather, which lets the organism (human) remain in homeostasis (a steady-state equilibrium). What this concretely means is that our 'shape' – especially the way our neurons are connected – is an embodiment of knowledge:

> Our ongoing life experience is a continuous stream of manifold forms of information which in turn patterns and repatterns configurations and processes within the brain [...]. This patterning is the physical embodiment of knowledge, which is the ongoing and ever-adapting patterned expectation against which we interpret the world of difference continuously thrown up by sense experience and mental manipulation. (Mobus and Kalton 2015, p. 85)

Some knowledge structures are evolved over multiple generations into a form of knowledge that affects the whole species. Moreover, agents throughout their lifetime develop in specific ways due to how their environment impact them. Notably, each agent learns specific things, having unique experiences in her life.

In summary, according to EST knowledge is closely tied to biology and evolution, amounting to a natural phenomenon that enables living organisms to avoid death. In simpler, more abundant, environments this can involve mostly hard-coded reflexes. In more complex environments, higher intelligence might be needed. Depending on context, knowledge needs to be 'good enough' for keeping the organism in homeostasis. Knowledge then is the internal structures that enable organisms to have a grasp of the world that mirrors reality to a high enough degree to sustain reproduction and survival.

5.3 Kornblith's Interpretation of Cognitive Ethology Revisited

We have seen how knowledge is characterized in cognitive psychology and EST. According to our interpretation, knowledge is seen as a natural phenomenon from both points of view – even though different levels of analysis and perspectives are in focus. And, as shown in the above quotes, 'knowledge' is in both cases used to denote the phenomenon in question. So where does this leave us? Can a plausible congruent cross-level series of theories of knowledge be identified? And (how) should our findings affect Kornblith's interpretation of knowledge, which is focused on cognitive ethology?

Now as stated in chapter 4, for two theories on different levels of analysis to be considered to be about the same phenomenon they should characterize their investigated phenomena in ways that are congruent, while also respecting our basic comprehension of said phenomena based on our rudimentary recognitional capacity and language competence.

Focusing on higher organisms, the EST view of knowledge as internal survival-beneficial structures is arguably congruent with the view(s) from cognitive psychology since the described memory systems amount to crucial structures that help us to stay alive. Admittedly, other internal structures also help in filling this role, but on the relevant level of analysis our memory systems are particularly important in guiding us so that we can tackle obstacles in our environment. Furthermore, both the EST and the cognitive psychological views of knowledge are arguably congruent with the view from cognitive ethology since, on this level and from this perspective, reliably produced true beliefs can be singled out as a particularly central form of internal survival-beneficial structures tied to particular memory systems. Finally, the three theories all use 'knowledge' to pinpoint what they are focusing on. In accordance with our approach, it is thus reasonable to consider that the theories are about the same natural phenomenon. This makes the claim '[k]nowledge, as it is portrayed in this literature, does causal and explanatory work.' (Kornblith 2002, pp. 28–29) not only applicable to cognitive ethology but also to EST and cognitive psychology.

Since cognitive epistemology draws on a richer set of scientific input than Kornblith's account does, it is expected that this approach might give us a conflicting account of some details. Next, this possibility will bear fruit in terms of three potential issues with Kornblith's interpretation of knowledge. It will be shown how these issues inform us that certain details need to be reinterpreted while it still remains plausible that the different theories all are about the same natural phenomenon.

The first issue with Kornblith's (2002) reliabilist account of knowledge is that it, arguably, most readily is compatible with how the reflexive memory processes (procedural and semantic memory) are characterized in cognitive psychology. Perceptual and motor processes (procedural memory), as well as our conceptual understanding (semantic memory), are easily characterized in externalist terminology, involving whether an agent has gotten her belief in a reliable way, through a reliable process, and whether she is favorably connected to the world (see, e.g., Pappas 2017; Parent 2017). But reflective memory processes (episodic memory, WM), on the other hand, might instead be more readily characterized in internalist terminology tied to reflection (Gärdenfors and Stephens 2017; Stephens 2019; Stephens and Tjøstheim 2020).

This is in itself not problematic, but when combined with Kornblith's (2002, 2012, 2019) interpretation, which explicitly downplays the importance of reflective processes for knowledge, it is:

> Reflection, by and large, does not provide for greater reliability. It does not, by and large, serve to guard against errors to which we would otherwise be susceptible. It does not, by and large, aid in the much needed project of cognitive self-improvement. It creates the illusion that it does all of these things, but it does not do any of them. (Kornblith 2012, p. 26)

Kornblith makes sure to highlight that reflection can be important in its own right. But he stresses that reflection typically does not provide added reliability – while '[f]rom an epistemological point of view, we should value reflection to the extent that, and only to the extent that, it contributes to our reliability.' (Kornblith 2012, p. 34).

The view from cognitive psychology instead encourages a more nuanced picture of what knowledge is by detailing what it involves and how it works. This means both a better understanding of the processes directly underlying Kornblith's reliabilist account of knowledge, as well as highlighting the importance of reflective processes, thereby indicating the plausibility of a complementary reflective form of knowledge to Kornblith's reliabilism. This would go against Kornblith's insistence that the only relevant account comes from cognitive ethology, as well as his interpretation that knowledge only comes in one form, instead fitting better with something more akin to, for example, Sosa's (see, e.g., 2007, 2009, 2010, 2011, 2015, 2017) division of knowledge into an animal form and a reflective form. Or at the very least indicate that a reflective component of knowledge should be more clearly acknowledged. This since the cognitive psychological view – specifically when focusing on higher organisms such as humans – deem reflective memory processes to be important in their own right, involving epistemic factors such as responsibility, rationality, and whether an agent has access to her beliefs through reflection. Notably, it is also possible to interpret reflective processes as actually being more reliable than Kornblith gives them credit for. Reflection often does add reliability through generalizability, flexibility, and creativity which can help higher organisms to handle new situations. This leaves the option of one form of knowledge – one which includes reflective aspects (Stephens and Tjøstheim 2020).

In later writings, Kornblith seemingly softens his tone, acknowledging a more influential role for reflection in line with the here advocated position:

> On less complicated issues, I don't need to stop to think things through. Once I see the evidence, I am immediately convinced; I know what to think. But this issue is complicated enough that I have to think through the evidence; I need to stop to reflect on what the evidence shows.
>
> Such situations are not altogether rare. They are an important part of our cognitive lives, and, at least on its face, it seems that some of our greatest epistemic achievements are a product of such self-conscious reflective thought. Even if much of our knowledge is easily attained, indeed, so easily attained that it requires no effort on our part at all, there is a good deal of knowledge as well which is a hard-won achievement, requiring careful reflection about just what we ought to believe. Any adequate survey of the phenomenon of knowledge must encompass both sorts of knowledge. (Kornblith 2021, p. 34)

The cognitive ethological view of knowledge as reliably produced true beliefs is compatible with the cognitive psychological view of knowledge as consisting in reflexive memory processes (while cognitive ethology can be interpreted as remaining fairly silent on the issue of reflective memory processes). It remains an open question if knowledge is best characterized as consisting in one, two, or three forms. Our claim is that all three sciences (cognitive psychology, EST, *and* cognitive ethology) would accept – indeed prefer – interpretations that acknowledge both the reflexive and the reflective aspects of knowledge (Gärdenfors and Stephens 2017; Stephens 2019; Stephens and Tjøstheim 2020).

The second issue with Kornblith's interpretation is found in how EST characterizes knowledge in a way that partly diverges from Kornblith's usage of 'truth' in his reliabilist account. According to EST all organisms have over evolutionary timescales been 'shaped' to function in particular ways that promote that they 'live on.' Such internal survival-beneficial structures *are* knowledge (Mobus and Kalton 2015). Knowledge is important, for living things, since it is what keeps them living. For higher organisms, living in complex environments, higher cognitive functions and beliefs are particularly central internal structures for keeping them alive. This much is seemingly compatible with how Kornblith (1993, p. 2) discusses how 'Quine has argued that knowledge of the world is possible, in part, because the world is divided by nature into kinds. At the same time, our psychological processes are so shaped by evolution as to be sensitive to those very natural kinds.' But, the *satisficing* aspect of the relation between organisms and the world is important (Simon 1956; Plotkin 1993, pp. 118, 171):

> [... I]t appears probable that, however adaptive the behavior of organisms in learning and choice situations, this adaptiveness falls far short of the ideal of "maximizing" postulated in economic theory. Evidently, organisms adapt well enough to "satisfice"; they do not, in general, "optimize." (Simon 1956, p. 129)

That a belief or action is satisficing, in this context, means that it must be good enough to satisfactory and sufficiently guarantee survival. This said, it is acceptable – indeed expected – that beliefs or actions occasionally turn out to be false or unsuccessful, in minor matters or very uncommon contexts/environments. That is, the focus on survival highlights how there is a real world, and that organisms would die if they did not meet the requirement of satisficing correspondence between their beliefs/actions and the world (see, e.g., Simon 1955, 1956; Millikan 1984; Gigerenzer 2001; Cellucci 2017; Artinger, Gigerenzer and Jacobs 2022). Kornblith is sensitive to this issue concerning fallibility and discusses a potential complication for the reliabilist position:

> What is being claimed here is that natural selection is selecting for knowledge-acquiring capacities, that is, processes of belief acquisition that tend to produce truths, and one might reasonably wonder whether this is the sort of thing for which natural selection might select. As many authors have argued, there are cases in which one process of belief acquisition is less reliable than another, and yet more conducive to survival. Faced with a choice between two such processes, natural selection will favor the more survival-conducive, and less truth-conducive, process. But then, it seems, it is conduciveness to survival that is being selected for rather than conduciveness to truth. (Kornblith 2002, p. 59)

Similar concerns are posed by other theorists (see, e.g., Stitch 1990; Plantinga 2011; Sage 2014; for defenses see, e.g., Deem 2018; Law 2012; see also, e.g., Dennett 1981; Feldman 1988). However, Kornblith argues that a biologically plausible interpretation offers a satisfying solution:

> This argument surely proves far too much, however, for it would show that conduciveness to survival is the only thing that is ever selected for. Were we to accept this argument, we would have to deny that the shape of a carnivore's teeth are selected for their ability to rip flesh, that the shape of the panda's thumb is selected for its ability to strip bamboo leaves from the stalk, and so on. This prohibition would fly in the face of current biological practice. Biologists do speak of these traits as selected for these particular functions, in spite of the fact that the practice of carrying out these functions can at times conflict with the goal of survival, just as the practice of acquiring true beliefs can at times conflict with the goal of survival. In spite of this, it is reasonable to claim that these traits are selected for these particular functions since the animals' abilities to carry out such functions do, on the whole, enhance survivability. (Kornblith 2002, p. 60)

We will follow Kornblith and view the overall circumstances as an ongoing balancing act that is taking place between the organism and its environment. So even though 'fitness-reliability' and 'truth-reliability' (Sage 2004) might come apart in certain situations, it is still fair to consider that the particular function of reliably generating true beliefs '[…] do, on the whole, enhance survivability.' (Kornblith 2002, p. 60). Boulter (2007) gives a stronger formulation:

> This serves to remind us of what ought to have been pretty obvious from the start, namely that too many false positives will be positively maladaptive since they will prevent the animal from engaging in other essential activities. Hiding in one's burrow all day in the false belief that a

> predator is lurking just outside may keep one alive in the short term. But neurotic animals do not feed well, nor do they tend to secure high quality mates (or any mates at all). (Boulter 2007, p. 377)

Nevertheless, the EST perspective can be seen to promote a slight specification or adjustment to Kornblith's reliabilist account. It might be imprecise to claim that true beliefs that are reliably produced amount to knowledge – specifically it is *satisficingly* true beliefs that are *satisficingly* reliably produced.[14]

Our claim here is, once more, that all three sciences would accept and prefer this interpretation.

Again, Kornblith in later writings seemingly comes very close to this conclusion himself although stating it in terms of 'nearly true,' 'approximately true,' 'roughly accurate,' or 'sufficiently reliable' (Kornblith 2021, pp. 132–133, 148):

> The world does, however, set a standard for knowledge. Remember that we were led to see knowledge as a scientific category because we need to appeal to this category in order to explain how creatures are able survive in a complex world. The environment a species inhabits creates certain informational demands; without an ability to pick up information about that environment, a species cannot endure. How reliable, then, must an animal's psychological processes be in picking up information about the world if those processes are to count as capable of producing knowledge? They must be reliable enough to allow the species to survive in that environment. This is not a standard which we have somehow imposed on the world because we care to have beliefs which are at least this reliably produced. It is a standard set by nature. Creatures with processes that meet such a standard will survive; those which don't, in W. V. Quine's memorable words, "have a pathetic but praiseworthy tendency to die before reproducing their kind" (1969, 126).
>
> We thus see that although the category of knowledge does involve meeting a certain standard— in particular, it involves having beliefs which are produced by processes which are *sufficiently* reliable— that standard— the level of reliability required— is not imposed by us; it is imposed by nature. Knowledge is thus properly viewed as a natural category, fully

[14] As long as this specification is understood as being implied, the shorter formulation might be sufficient. See also Cellucci (2017) who makes a similar point but argues that 'plausibility' ought to replace the truth-condition.

> amenable to scientific investigation. (Kornblith 2021, p. 148, italics added)

Even though this seems compatible with the argumentation here presented, Kornblith does not address any possible consequence for the reliabilist knowledge-definition.

The two previous issues can be seen as direct criticisms of Kornblith's account. The third issue, stemming from the dynamical perspective of EST, can be interpreted in this way as well but we will primarily see it as encouraging future research that looks closer at what it means to see knowledge as a natural *process*.

Kornblith (2002) takes knowledge to be a natural phenomenon and a natural *kind*. He considers '[...] natural kinds to be homeostatically clustered properties, properties that are mutually supporting and reinforcing in the face of external change.' (Kornblith 2002, p. 61). This means that Kornblith holds knowledge to have a certain well-behaved stability to it (Kornblith 2002, p. 62). In other words, natural kinds consist of certain properties that cluster together in ways that are stable enough to tackle external pressure. In this way they form well-behaved categories.

Kornblith's perspective is plausible but, as pointed out by various process-oriented philosophers, humans are the product of natural biological *processes* where knowledge plays a central role (Whitehead 1925, 1929; Laszlo 1972a, 1972b; Plotkin 1993; Cellucci 2017; Nicholson and Dupré 2018). For example, Plotkin (1993) describes knowledge as being equivalent to adaptation, where '[...] adaptations are biological knowledge, and knowledge as we commonly understand the word is a special case of biological knowledge.' (Plotkin 1993, p. xv). Cellucci (2017) takes knowledge to be a natural phenomenon, where knowledge is '[...] not merely a state of mind, but rather a response to the environment that is essential for survival. [... K]nowledge is a natural process, continuous with the biological processes by which life is sustained and evolved, and has a vital role, in the literal sense that life exists only insofar as there is knowledge.' (Cellucci 2017, p. 65). Nicholson and Dupré (2018) argue that '[t]he reason why mechanistic explanations provide insights (to the extent that they do) is that the components of the mechanisms being described are sufficiently stable on the timescale of the phenomena under investigation.' (Dupré and Nicholson 2018, p. 29).

What there *is*, on these accounts, are thus various natural *processes* that are stable to varying degrees. Admittedly, this stability can motivate talk about 'states' and ‘kinds,’ which would be compatible with Kornblith’s view, but nevertheless the issue deserves more attention.

Arguably, this situation indicates that a pluralistic stance regarding the classification of natural phenomena (such as knowledge) is plausible, since there is no obviously ‘correct’ way to cut nature at its joints – which was highlighted in our previous discussion of the first issue concerning how memory processes should be seen and classified (Dupré and Nicholson 2018, p. 23; see also Dupré 1993; Stephens 2016; Stephens et al. 2021).[15] All three sciences should be able to accept this interpretation.

By heeding the cognitive epistemological framework, looking to different sciences, it has been possible to sketch the outlines of a multi-perspective account of the natural phenomenon of knowledge. It has been argued that it is reasonable to think that the EST view of knowledge as internal survival-beneficial structures and the cognitive psychological view of higher organisms’ knowledge as consisting in reflexive and reflective memory processes are congruent with each other, as well as with the reliabilist account from cognitive ethology. Furthermore, it is reasonable to think that they concern the same natural phenomenon. Notably, we have been able to pinpoint three issues with Kornblith’s interpretation of the knowledge-account from cognitive ethology. First, knowledge can plausibly be seen to consist in one, two, or three, forms. Second, higher organisms’ knowledge should be seen to involve *satisficingly* true beliefs that are *satisficingly* reliably produced. Third, the dynamical perspective highlights the need for further exploration concerning the processual nature of the natural phenomenon of knowledge.

[15] Here Kusch’s (2005) discussion of the possibility of choosing to focus on the sociology of scientific knowledge is relevant. From this scientific perspective, knowledge might, arguably, be seen as a *social* kind. In his reply to Kusch, Kornblith (2005) does not address this issue head-on (but see Kornblith 2002, pp. 70–102 for a discussion of knowledge and social practices in connection to, for example, Davidson 1984, 1999; Brandom 1994; and Williams 2000). According to cognitive epistemology, sociologists of scientific knowledge are perfectly within their rights to investigate knowledge as a social kind (or however they choose – which is for science, not philosophy, to decide). If we stipulate that there is such a social kind doing ‘causal and explanatory work’ (Kornblith 2002, pp. 28–29) in the sociology of scientific knowledge research, cognitive epistemology acknowledges this and encourages (as is discussed in chapter 4) an exploration of whether this account is congruent with other sciences’ accounts and whether it is reasonable to think that they concern the same natural phenomenon.

Chapter 6

Concluding Remarks

WE HAVE ADDRESSED the question 'What is knowledge?' by looking at intuition-based epistemology, then naturalistic approaches, and finally one particular naturalistic approach: cognitive epistemology.

Intuition-based epistemology was found to characteristically focus on how 'knowledge' is used linguistically or conceptually rather than on what knowledge *is*. The usage of intuitions as evidence was considered to be problematic since even though intuitions concerning knowledge can pick out instances of the phenomenon, this is not always the case. Moreover, it was argued that since experimental results indicate that people's intuitions vary a great deal, while no conclusive systematicity can be found, intuition-based approaches cannot provide a solid foundation to answer the initial question.

Different naturalistic approaches were then discussed, where it is knowledge the natural phenomenon – not the concept – that is in focus. It was maintained that most philosophers nowadays accept ontological naturalism but that several different methodological approaches are followed – the common thread being that science should decidedly influence philosophy.

Grounded in naturalism, a specific cognitive epistemological approach was presented, accepting ontological naturalism, methodological cooperative naturalism, and evolutionary epistemology. This approach followed

Kornblith (2002) in looking to science to understand the natural phenomenon of knowledge. However, given its pluralistic commitment, cognitive epistemology let us focus on cognitive psychology and EST, as a complement to Kornblith's sole focus on cognitive ethology. The emerging picture indicated that some refinements to Kornblith's approach and knowledge-account were necessary.

By triangulating knowledge in this way, we arrived at a view where higher organisms' knowledge can reasonably be characterized as involving either one, two, or three distinct forms. From the perspectives of cognitive psychology, EST, and cognitive ethology, knowledge is a natural phenomenon amounting to internal survival-beneficial structures, which for higher organisms importantly involve reflexive and reflective memory processes that satisficingly reliably produce satisficingly true beliefs.

Cognitive epistemology thus lets us arrive at an answer to our initial question. The significance of this result can be seen if we broaden our outlook. In addition to laying the foundation for a better understanding of what knowledge is, the approach here pursued opens for, at least, four natural continuations. First, as already mentioned, the dynamical perspective highlights the need for further exploration concerning the processual nature of the natural phenomenon of knowledge. Second, this thesis has focused on two sciences' view of knowledge – in addition to the one Kornblith focuses on. A sensible next step, to further solve the Rubik's Cube of what knowledge is in greater detail, would be to involve the point of view of even more sciences. This exploration could, for example, go 'downwards' towards neuroscience or 'upwards' towards social psychology. Third, the here developed approach and knowledge-account can be used to address epistemological 'problems' (as is done in Stephens et al. 2021). This means that it might be fruitful for future research to explore how cognitive epistemology can offer dissolving insights to influential puzzles that have long preoccupied epistemology. Fourth, further exploration could also include a generalization of the approach to address metaphysics and philosophy of mind. In this way the same method that was applied to the investigation of what knowledge is could be applied to finding out the fundamental nature of reality and the mind.

6.1 Scientific Publications

The papers that are included in this thesis all try to enhance our understanding of the natural phenomenon of knowledge. They were, however, not written with a specific overarching argument in mind. It is my hope that the above discussion nevertheless has presented a coherent case. Paper I deals with methodological issues in naturalistic epistemology, arguing the case for a pluralistic stance. This provided the foundation for the cognitive epistemological approach in general and the discussion in chapter 4 specifically. Papers II–IV argue that a pluralistic perspective lets us develop a multi-level account of knowledge seen as a natural phenomenon, by focusing first on LTM and then on WM. The papers together form a detailing of what the natural phenomenon of knowledge amounts to, with a special focus on a cognitive psychological perspective. Chapter 5, and especially section 5.1, can be seen as a condensed presentation of these discussions. Finally, Paper V shows how a dynamical perspective might have a dissolving influence on a traditional epistemological 'problem.' This perspective showcased the problem-dissolving possibility of the cognitive epistemological approach while also underscoring the dynamical perspective that was highlighted in section 5.2. The three co-written papers' (II, IV, and V) division of labor will be presented.

Paper I – '**A pluralist account of knowledge as a natural kind**' (Andreas Stephens 2016) – presents Kornblith's account of knowledge as a natural kind. After discussing various aspects of Kornblith's account, a central methodological issue is identified. Kornblith's explicitly promoted sole reliance on cognitive ethology is questioned. After highlighting the fact that other sciences, such as cognitive neuroscience and cognitive psychology, also investigate the natural phenomenon knowledge it is argued that a more pluralistic stance than that which Kornblith presents is called for. Finally, it is argued that Kornblith's theory, which has many benefits, can be recast in a pluralistic form which ought to be seen as a more fruitful naturalistic alternative. While this methodological discussion – highlighting the issue of pluralism – focuses on how Kornblith's account can be revised, it lays the groundwork for the cognitive epistemological approach outlined in chapter 4.

Paper II – '**Induction and knowledge-what**' (Peter Gärdenfors and Andreas Stephens 2017 [2018]) – argues that the two commonly recognized knowledge forms, knowledge-how and knowledge-that, should be accompanied by a third form: *knowledge-what*. Knowledge-what concerns relations between properties and categories and it is argued that it cannot be reduced to knowledge-that. This tripartite partitioning of knowledge is supported by mapping it onto the LTM systems: procedural-, semantic- and episodic memory. It is further argued that the role of inductive reasoning is to generate knowledge-what. Conceptual spaces are used to model knowledge-what and the relations between properties and categories involved in induction. This cognitive psychological discussion is relevant for chapter 5 – specifically sections 5.1 and 5.3. While the paper specifically argues for a tripartite view of knowledge, chapter 5 instead takes a more noncommittal meta-perspective.

This paper grew out of Stephens' master's thesis in cognitive science. Stephens developed the evolutionary epistemological account connecting knowledge to adaptation optimization. Gärdenfors came up with the idea of connecting induction and semantic memory to conceptual knowledge. Both Stephens and Gärdenfors were involved throughout the process of rewriting Stephens' thesis into its present article form.

Paper III – '**Three levels of naturalistic knowledge**' (Andreas Stephens 2019) – focuses on LTM, striving to lay the groundwork for an integration of knowledge-accounts from different levels of analysis. It is found that procedural knowledge-how (perceptual- and motor pathways/procedural memory) and conceptual knowledge-what (associative pathways/semantic memory), on lower neuroscientific and psychological levels of analysis, are congruent with a higher-level ethological reliabilist account. This account is also congruent with System 1 from dual process theory on a higher psychological level. Moreover, it is argued that the inclusion of knowledge-that (attentional- and executive pathways/episodic memory), on a lower level of analysis can account for System 2 on the psychological level. This discussion is relevant for chapters 4 and 5 since it highlights the multi-level approach to the natural phenomenon under investigation while also providing a detailed outline of relevant cognitive psychological underpinnings.

Paper IV – '**The cognitive philosophy of reflection**' (Andreas Stephens and Trond A. Tjøstheim 2020 [2022]) – questions Kornblith's argument that many traditional philosophical accounts involve problematic views of reflection (understood as second-order mental states). According to Kornblith, reflection does not add reliability, which makes it unfit to underlie a separate form of knowledge. Focusing on WM, it is shown that a broader understanding of reflection, encompassing Type 2 processes (System 2), WM, and episodic LTM, can provide philosophy with elucidating input that a restricted view misses. It is further argued that reflection in fact often does add reliability, through generalizability, flexibility, and creativity that is helpful in newly encountered situations, even if the restricted sense of both reflection and knowledge is accepted. And so, a division of knowledge into one reflexive form and one reflective form remains a plausible option. This cognitive psychological discussion broadens the input for chapter 5, specifically concerning the role of WM.

Both authors contributed equally to the general idea. Stephens took lead and developed the overall structure of the paper, as well as sections 1–3. Tjøstheim developed section 4. Both authors contributed to each other's sections.

Paper V – 'A dynamical perspective on the generality problem**'** (Andreas Stephens, Trond A. Tjøstheim, Maximilian K. Roszko, and Erik J. Olsson 2021) – addresses the generality problem, which is commonly considered to be a critical difficulty for reliabilism. This paper presents a dynamical and process-oriented perspective, in line with EST, on the problem in the spirit of naturalized epistemology. According to this outlook, it is worth investigating how token belief-forming processes instantiate specific types in the biological agent's cognitive architecture (including other relevant embodied features) and background experience, consisting in the process of attractor-guided neural activation. While the discussion of the generality problem assigns 'scientific types' to token processes, it represents a unified account in the sense that it incorporates contextual and common-sense features emphasized by other authors. The paper which addresses a specific epistemological 'problem' can be viewed as a case study, indicating that cognitive epistemology might be used to fruitfully address other epistemological 'problems.' This discussion is relevant for chapter 5 – specifically sections 5.2 and 5.3 – since it amounts to an application of the cognitive epistemological approach to a well-known problem in epistemology and offers a novel perspective which has a dissolving potential.

All authors contributed to the general idea of the paper. Stephens and Olsson took lead in the philosophical discussion in sections 1 and 2, with input from Tjøstheim and Roszko. Stephens wrote-up a first draft of this section which was discussed and edited by the other authors. Stephens, Tjøstheim, and Roszko took lead in the cognitive scientific discussion in sections 3–5, with input from Olsson. Stephens and Tjøstheim wrote-up a first draft of these sections which was discussed and edited by the other authors. Stephens and Olsson took lead in the discussion in section 6, with input from Tjøstheim and Roszko. Stephens and Olsson wrote-up a first draft of this section which was discussed and edited by the other authors.

References

Alexander, J. (2012). *Experimental philosophy: An introduction*. Cambridge: Polity.

Alexander, J., and Weinberg, J. M. (2014). The "unreliability" of epistemic intuitions. In E. Machery and E. O'Neill (eds.), *Current controversies in experimental philosophy* (pp. 128-145). New York: Routledge.

Angere, S. (2010). *Theory and reality: Metaphysics as second science*. Doctoral dissertation, Department of Philosophy, Lund University.

Artinger, F. M., Gigerenzer, G., and Jacobs, P. (2022). Satisficing: Integrating two traditions. *Journal of Economic Literature*, *60*(2), 598-635.

Ashby, W. R. (1956). *An introduction to cybernetics*. London: Chapman and Hall.

Badcock, P. B. (2012). Evolutionary systems theory: A unifying meta-theory of psychological science. *Review of General Psychology*, *16*(1), 10-23.

Badcock, P. B., Friston, K. J., and Ramstead, M. J. (2019). The hierarchically mechanistic mind: A free-energy formulation of the human psyche. *Physics of life Reviews*, *31*, 104-121.

Badcock, P. B., Friston, K. J., Ramstead, M. J., Ploeger, A., and Hohwy, J. (2019). The hierarchically mechanistic mind: An evolutionary systems theory of the human brain, cognition, and behavior. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(6), 1319-1351.

Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.

Bealer, G. (1992). The incoherence of empiricism. *Proceedings of the Aristotelian Society* (Supplementary volume), *66*, 99-143.

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior*, *81*(2), 179-209.

Boulter, S. J. (2007). The "evolutionary argument" and the metaphilosophy of commonsense. *Biology and Philosophy*, *22*(3), 369-382.

Bradie, M. and Harms, W. (2020). Evolutionary epistemology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 edition), URL = <https://plato.stanford.edu/archives/spr2020/entries/epistemology-evolutionary/>.

Brandom, R. (1994). *Making it explicit: Reasoning, representing and discursive commitment*. Cambridge, Mass.: Harvard University Press.

Buckwalter, W. (2016). Intuition fail: Philosophical activity and the limits of expertise. *Philosophy and Phenomenological Research*, *92*(2), 378-410.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (ed.), *The philosophy of Karl R. Popper* (pp. 412-463). LaSalle, IL: Open Court.

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews*, *9*(3), 399-431.

Cappelen, H. (2012). *Philosophy without intuitions*. Oxford: Oxford University Press.

Carpenter, R. H. S. (2004). Homeostasis: A plea for a unified approach. *Advances in Physiology Education*, *28*(4), 180-187.

Cellucci, C. (2014). Rethinking philosophy. *Philosophia*, *42*(2), 271-288.

Cellucci, C. (2017). *Rethinking knowledge: The heuristic view* (European Studies in Philosophy of Science Vol. 4). Cham: Springer.

Cellucci, C. (2022). *The making of mathematics: Heuristic philosophy of mathematics*. Cham: Springer.

Chalmers, D. J. (2014). Intuitions in philosophy: A minimal defense. *Philosophical Studies*, *171*(3), 535-544.

Churchland, P. S. (1982). Mind-brain reduction: New light from the philosophy of science. *Neuroscience*, *7*(5), 1041-1047.

Cohen, N. J., and Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 201-210.

Cohnitz, D. (2020). Thought experiments and the (ir-)relevance of intuitions in philosophy. In Hermann, J., Hopster, J., Kalf, W., and Klenk, M. (eds.), *Philosophy in the age of science?: Inquiries into philosophical progress, method, and societal relevance* (II: 5). Lanham, Maryland: Rowman & Littlefield Publishers.

Comesaña, J., and Klein, P. (2019). Skepticism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 edition), URL = <https://plato.stanford.edu/archives/win2019/entries/skepticism/>.

Corcoran, A. W. and Hohwy, J. (2018). Allostasis, interoception, and the free energy principle: Feeling our way forward. In M. Tsakiris and H. De Preester (eds.), *The interoceptive mind: From homeostasis to awareness* (pp. 272–292). Oxford: Oxford University Press.

Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Oxford University Press.

Davidson, D. (1999). Reply to Simon J. Evnine. In L. E. Hahn (ed.), *The philosophy of Donald Davidson* (pp. 305–310). Chicago: Open Court.

Davidson, M. (1983). *Uncommon sense: The life and thought of Ludwig von Bertalanffy, father of general systems theory*. Los Angeles: Tarcher.

Day, T. A. (2005). Defining stress as a prelude to mapping its neurocircuitry: No help from allostasis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *29*(8), 1195-1200.

Deem, M. J. (2018). A flaw in the Stich–Plantinga challenge to evolutionary reliabilism. *Analysis*, *78*(2), 216-225.

Dennett, D. C. (1981). Making sense of ourselves. *Philosophical Topics*, *12*(1), 63-81.

Dere, E., Kart-Teke, E., Huston, J. P., and Silva, M. D. S. (2006). The case for episodic memory in animals. *Neuroscience & Biobehavioral Reviews*, *30*(8), 1206-1224.

Deutsch, M. (2015). *The myth of the intuitive: Experimental philosophy and philosophical method*. Cambridge, MA: MIT Press.

Devitt, M. (2015). Relying on intuitions: Where Cappelen and Deutsch go wrong. *Inquiry*, *58*(7-8), 669-699.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge, MA: Harvard University Press.

Dupré, J., and Nicholson, D. J. (2018). A manifesto for a processual philosophy of biology. In Nicholson, D. J., and Dupré, J. (eds.) (2018). *Everything flows: Towards a processual philosophy of biology* (pp. 3-45). Oxford: Oxford University Press.

Feldman, R. (1988). Rationality, reliability, and natural selection. *Philosophy of Science*, *55*(2), 218-227.

Feldman, R. (2012). Epistemology naturalized. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Fall 2012 edition), URL = <https://plato.stanford.edu/archives/sum2012/entries/epistemology-naturalized/>.

Feyerabend, P. K. (1975). *Against method: Outline of an anarchistic theory of knowledge*. London: New Left Books.

Fischer, E., and Sytsma, J. (2021). Zombie intuitions. *Cognition*, *215*, 104807.

Friston, K. (2009). The free-energy principle: A rough guide to the brain?. *Trends in cognitive sciences*, *13*(7), 293-301.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127-138.

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1-3), 70-87.

Gärdenfors, P., and Stephens, A. (2017 [2018]). Induction and knowledge-what. *European Journal for Philosophy of Science*, *8*(3), 471-491.

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2019). *Cognitive neuroscience: The biology of the brain* (5th edition). New York: W. W. Norton & Company.

Gigerenzer, G. (2001). The adaptive toolbox In G. Gigerenzer and R. Selten (eds.), *Bounded rationality: The adaptive toolbox* (pp. 37-50). Cambridge, MA: MIT Press.

Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science*. Chicago: The University of Chicago Press.

Goldman, A. I. (1979). What is justified belief?. In G. Pappas (ed.), *Justification and Knowledge: New Studies in Epistemology* (1-23). Dordrecht, Reidel.

Goldman, A. I. (2005). Kornblith's naturalistic epistemology. *Philosophy and Phenomenological Research*, *71*(2), 403-410.

Goldman, A. I. (2007). Philosophical intuitions: Their target, their source and their epistemic status. *Grazer Philosophische Studien*, *4*, 1-26.

Graf, P., and Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, memory, and cognition*, *11*(3), 501-518.

Griffin, D. R. (1976). *The question of animal awareness*. New York: Rockefeller University Press.

Griffin, D. R. (1978). Prospects for a cognitive ethology. *The Behavioral and Brain Sciences*, *4*, 527-538.

Hawking, S. and Mlodinow, L. (2011). *The grand design*. London: Bantam Books.

Hofkirchner, W. (2005). Ludwig von Bertalanffy, forerunner of evolutionary systems theory. In *The new role of systems sciences for a knowledge-based society: Proceedings of the first world congress of the International Federation for Systems Research (Vol. 6)*. Kobe, Japan.

Ichikawa, J. J., and Steup, M. (2018). The analysis of knowledge. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 edition), URL = <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>.

Itakura, H. (2001). Describing conversational dominance. *Journal of Pragmatics*, *33*(12), 1859-1880.

Jackson, F. (2011). On Gettier holdouts. *Mind and Language*, *26*(4), 468-481.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (eds.) (2013). *Principles of neural science* (5th edition). New York: McGraw-Hill, Health Professions Division.

Kim, M., and Yuan, Y. (2015). No cross-cultural differences in the Gettier car case intuition: A replication study of Weinberg et al. 2001. *Episteme*, *12*(3), 355-361.

Kingstone, A., Smilek, D., and Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317-340.

Kitcher, P. (1992). The naturalists return. *The Philosophical Review*, *101*(1), p. 53-114.

Knobe, J., and Nichols, S. (ed.) (2008). *Experimental philosophy*. Oxford: Oxford University Press.

Knobe, J. and Nichols, S. (2017). Experimental philosophy. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 edition), URL = <https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/>.

Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge, MA: MIT Press.

Kornblith, H. (1995). Naturalistic epistemology and its critics. *Philosophical Topics*, *23*(1), 237-255.

Kornblith, H. (1999). In defense of a naturalized epistemology. In J. Greco and E. Sosa (eds.), *The Blackwell Guide to Epistemology* (pp. 158-169). Malden, MA: Blackwell.

Kornblith, H. (2002). *Knowledge and its place in nature*. New York: Oxford University Press.

Kornblith, H. (2005). Replies to Alvin Goldman, Martin Kusch and William Talbott. *Philosophy and Phenomenological Research*, *71*(2), 427-441.

Kornblith, H. (2012). *On reflection*. Oxford: Oxford University Press.

Kornblith, H. (2019). *Second thoughts and the epistemological enterprise*. Cambridge: Cambridge University Press.

Kornblith, H. (2021). *Scientific epistemology: An introduction*. Oxford: Oxford University Press.

Kruglanski, A. W., Jasko, K., and Friston, K. J. (2020). All thinking is 'wishful' thinking. *Trends in Cognitive Sciences*, *24*(6), 413-424.

Kusch, M. (2005). Beliefs, kinds and rules: A comment on Kornblith's *Knowledge and its place in nature*. *Philosophy and Phenomenological Research*, *71*(2), 411-419.

Laird, J. E., Lebiere, C., Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*(4), 13-26.

Laszlo, E. (1972a). *The systems view of the world: The natural philosophy of the new developments in the sciences*. Oxford: Basil Blackwell.

Laszlo, E. (1972b). *Introduction to systems philosophy: Toward a new paradigm of contemporary thought*. New York: Gordon and Breach.

Law, S. (2012). Naturalism, evolution and true belief. *Analysis*, *72*(1), 41-48.

Lin, H. (2023). Bayesian epistemology. In E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023 edition),
URL = <https://plato.stanford.edu/archives/win2023/entries/epistemology-bayesian/>.

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford: Oxford University Press.

Margolis, E., and Laurence, S. (2021). Concepts. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 edition),
URL = <https://plato.stanford.edu/archives/spr2021/entries/concepts/>.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

McEwen, B. S., and Wingfield, J. C. (2010). What's in a name?: Integrating homeostasis, allostasis and stress. *Hormones and Behavior*, *57*(2), 105-111.

Meunier, D., Lambiotte, R., and Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, *4*, 200.

Millikan, R. G. (1984). Naturalist reflections on knowledge. *Pacific Philosophical Quarterly*, *65*(4), 315-334.

Mitchell, S. D. (2002). Integrative pluralism. *Biology and Philosophy*, *17*(1), 55-70.

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Mobus, G. E. (2022). *Systems science: Theory, analysis, modeling, and design*. Cham: Springer.

Mobus, G. E., and Kalton, M. C. (2015). *Principles of systems science*. New York: Springer.

Nagel, J. (2012). Intuitions and experiments: A defense of the case method in epistemology. *Philosophy and Phenomenological Research*, *85*(3), 495-527.

Nichols, S. (2004). Folk concepts and intuitions: From philosophy to cognitive science. *Trends in Cognitive Sciences*, *8*(11), 514-518.

Nicholson, D. J., and Dupré, J. (eds.) (2018). *Everything flows: Towards a processual philosophy of biology*. Oxford: Oxford University Press.

Olsson, E. J. (2021). Explicationist epistemology and the explanatory role of knowledge. *Journal for General Philosophy of Science*, 1-20.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and cognition*, *14*(1), 30-80.

Papineau, D. (2021a). The disvalue of knowledge. *Synthese*, *198*(6), 5311-5332.

Papineau, D. (2021b). Naturalism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 edition),
URL = <https://plato.stanford.edu/archives/sum2021/entries/naturalism/>.

Pappas, G. (2017). Internalist vs. externalist conceptions of epistemic justification. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 edition),
URL = <https://plato.stanford.edu/archives/fall2017/entries/justep-intext/>.

Parent, T. (2017). Externalism and self-knowledge. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 edition),
URL = <https://plato.stanford.edu/archives/fall2017/entries/self-knowledge-externalism/>.

Plantinga, A. (2011). Content and natural selection. *Philosophy and Phenomenological Research*, *83*(2), 435-458.

Plotkin, H. (1993). *Darwin machines and the nature of knowledge*. Cambridge, MA: Harvard University Press.

Pust, J. (2019). Intuition. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 edition),
URL = <https://plato.stanford.edu/archives/sum2019/entries/intuition/>.

Quine, W. V. O. (1969a). Epistemology naturalized. In *Ontological relativity and other essays* (pp. 69-90). New York: Columbia University Press.

Quine, W. V. O. (1969b). Natural kinds. In *Ontological relativity and other essays* (pp. 114-138). New York: Columbia University Press.

Quinlan, P., and Dyson, B. (2008). *Cognitive psychology*. Harlow: Pearson Education Limited.

Ramstead, M. J. D., Badcock, P. B., and Friston, K. J. (2018a). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, *24*, 1-16.

Ramstead, M. J., Badcock, P. B., and Friston, K. J. (2018b). Variational neuroethology: Answering further questions: Reply to comments on Answering Schrödinger's question: A free–energy formulation. *Physics of Life Reviews*, *24*, 59-66.

Sage, J. (2004). Truth-reliability and the evolution of human cognitive faculties. *Philosophical Studies*, *117*(1/2), 95-106.

Sayre, K. M. (1976). *Cybernetics and the philosophy of mind.* London: Routledge & Kegan Paul Ltd.

Schrödinger, E. (1944). *What is life?: The physical aspect of the living cell.* Cambridge: Cambridge University Press.

Sellars, R. W. (1919). The epistemology of evolutionary naturalism. *Mind*, *28*(112), 407-426.

Seyedsayamdost, H. (2015a). On gender and philosophical intuition: Failure of replication and other negative results. *Philosophical Psychology*, *28*(5), 642-673.

Seyedsayamdost, H. (2015b). On normativity and epistemic intuitions: Failure of replication. *Episteme*, *12*(1), 95-116.

Silva, A. J., and Bickle, J. (2009). The science of research and the search for molecular mechanisms of cognitive functions. In J. Bickle (ed.), *The Oxford handbook of philosophy and neuroscience* (pp. 92-126). New York: Oxford University Press.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129-138.

Sorensen, R. (2020). Epistemic paradoxes. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 edition), URL = <https://plato.stanford.edu/archives/fall2020/entries/epistemic-paradoxes/>.

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge (Vol. I).* Oxford: Oxford University Press.

Sosa, E. (2009). *Reflective knowledge: Apt belief and reflective knowledge (Vol. II).* Oxford: Oxford University Press.

Sosa, E. (2010). How competence matters in epistemology. *Philosophical Perspectives*, *24*(1), 465-475.

Sosa, E. (2011). *Knowing full well.* Princeton, NJ: Princeton University Press.

Sosa, E. (2015). *Judgment and agency*. Oxford: Oxford University Press.

Sosa, E. (2017). *Epistemology*. Princeton: Princeton University Press.

Starmans, C., and Friedman, O. (2012). The folk conception of knowledge. *Cognition*, *124*(3), 272-283.

Steine-Hanson, Z., Koh, N., and Stocco, A. (2018). Refining the common model of cognition through large neuroscience data. *Procedia Computer Science*, *145*, 813-820.

Stephens, A. (2016). A pluralist account of knowledge as a natural kind. *Philosophia*, *44*(3), 885-903.

Stephens, A. (2019). Three levels of naturalistic knowledge. In M. Kaipainen, F. Zenker, A. Hautamäki, and P. Gärdenfors (eds.), *Conceptual spaces: Elaborations and applications* (Synthese Library, Studies in Epistemology, Logic, Methodology, and Philosophy of Science, Vol. 405, pp. 57-73). Cham: Springer Nature Switzerland.

Stephens, A., and Tjøstheim, T. A. (2020 [2022]). The cognitive philosophy of reflection. *Erkenntnis*, *87*, 2219-2242.

Stephens, A., Tjøstheim, T. A., Roszko, M., and Olsson, E. J. (2021). A dynamical perspective on the generality problem. *Acta Analytica*, *36*(3), 409-422.

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, *106*(1), 5-15.

Sterling, P., and Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In S. Fisher and J. T. Reason (eds.), *Handbook of life stress, cognition, and health* (pp. 629-649). Chicester, NY: Wiley.

Stich, S. P. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge, MA: MIT Press.

Stocco, A., Laird, J., Lebiere, C., and Rosenbloom, P. (2018). Empirical evidence from neuroimaging data for a Standard Model of the Mind. In C. Kalish, M. Rau, J. Zhou, and T. T. Rogers (eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 1094-1099).

Stocco, A., Sibert, C., Steine-Hanson, Z., Koh, N., Laird, J. E., Lebiere, C. J., and Rosenbloom, P. (2021). Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains. *NeuroImage*, *235*, 118035.

Talbot, B. (2009). Psychology and the use of intuitions in philosophy. In D. Cohnitz and S. Häggqvist (eds.), *Studia Philosophica Estonica 2:2*, 157-176. (Special issue: *The role of intuitions in philosophical methodology.*)

Talbott, W. (2005). Universal knowledge. *Philosophy & Phenomenological Research*, *71*(2), 420-426.

Templer, V. L., and Hampton, R. R. (2013). Episodic memory in nonhuman animals. *Current Biology*, *23*(17), R801-R806.

Ten Berge, T., and Van Hezewijk, R. (1999). Procedural and declarative knowledge: An evolutionary perspective. *Theory & Psychology*, *9*(5), 605-624.

Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie (Journal of Comparative Ethology)*, *20*(4), 410-433.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1-12.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1-25.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human. In H. Terrance and J. Metcalfe (eds.), *The missing link in cognition: Origins of self-reflective consciousness* (3-56). New York: Oxford University Press.

Turchin, V. F. (1993). On cybernetic epistemology. *Systems Research*, *10*(1), 3-28.

Turri, J. (2013). A conspicuous art: Putting Gettier to the test. *Philosopher's Imprint*, *13*(10), 1-16.

Turri, J. (2016). Knowledge judgments in "Gettier" cases. In J. Sytsma and W. Buckwalter (eds.), *A companion to experimental philosophy* (pp. 337-348). Malden, Mass.: Wiley-Blackwell.

von Bertalanffy, L. (1968). *General system theory.* New York: George Braziller.

Ward, J. (2010). *The student's guide to cognitive neuroscience.* New York: Psychology Press.

Whitehead, A. N. (1925). *Science and the modern world.* Cambridge: Cambridge University Press.

Whitehead, A. N. (1929). *Process and reality: An essay in cosmology*. New York: The Free Press.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine.* Cambridge, MA: MIT Press.

Williams, M. (2000). Dretske on epistemic entitlement. *Philosophy and Phenomenological Research, 60*(3), 607-612.

Williamson, T. (2000). *Knowledge and its limits*, Oxford: Oxford University Press.

Williamson, T. (2007). *The philosophy of philosophy*. Malden, MA: Blackwell Publishing.

Williamson, T. (2011). Philosophical expertise and the burden of proof. *Metaphilosophy*, *42*(3), 215-229.

Yee, E., Chrysikou, E. G., and Thompson-Schill, S. L. (2014). Semantic memory. In K. Ochsner and S. Kosslyn (eds.), *The Oxford handbook of cognitive neuroscience: Volume 1, core topics (pp. 353-374).* Oxford: Oxford University Press.

Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge: Cambridge University Press.

Papers

Paper I

A. Stephens
A pluralist account of knowledge as a natural kind
Philosophia, 2016, *44*(3), 885-903

Philosophia (2016) 44:885–903
DOI 10.1007/s11406-016-9738-3

A Pluralist Account of Knowledge as a Natural Kind

Andreas Stephens[1]

Received: 21 March 2016 / Revised: 6 June 2016 / Accepted: 16 June 2016 /
Published online: 30 June 2016

Abstract In an attempt to address some long-standing issues of epistemology, Hilary Kornblith proposes that knowledge is a natural kind the identification of which is the unique responsibility of one particular science: cognitive ethology. As Kornblith sees it, the natural kind thus picked out is knowledge as construed by reliabilism. Yet the claim that cognitive ethology has this special role has not convinced all critics. The present article argues that knowledge plays a causal and explanatory role within many of our more fruitful current theories, diverging from the reliabilist conception even in disciplines that are closely related to cognitive ethology, and thus still dealing with knowledge as a natural as opposed to a social phenomenon, where special attention will be given to cognitive neuroscience. However, rather than discarding the natural kind approach altogether, it is argued that many of Kornblith's insights can in fact be preserved within a framework that is both naturalist and pluralist.

Keywords Hilary Kornblith · Knowledge · Naturalistic epistemology · Cognitive neuroscience · Cognitive ethology · Pluralism · Natural kind

Introduction

Knowledge is important to us both in our daily life and in science. However, philosophical investigations and discussions regarding how we ought to view knowledge have been going on for millennia without clear results, often following a historical split between those focusing on internal aspects, such as Bonjour (1985) and Chisholm (1988), or external aspects, such as Dretske (1981) and Goldman (1986). Similarly, many philosophers nowadays would consider themselves to be heeding some form of naturalism, although finding a generally approved naturalistic approach is difficult. Naturalism has instead been interpreted and promoted in many different

✉ Andreas Stephens
andreasstephens@gmail.com

[1] Lund University, Lund, Sweden

forms, from the pragmatism of Peirce (1877), James (1907) and Dewey (1938) to the eliminative materialism of Dennett (1996) and the Churchlands (1998).

In *Knowledge and its Place in Nature* (2002) Hilary Kornblith presents a naturalistic epistemological theory, based on cognitive ethology, according to which knowledge should be seen as a natural phenomenon and a natural kind requiring reliably produced true belief. I view Kornblith's theory as a promising candidate for a fruitful naturalistic epistemology, but his choice to use cognitive ethology as his sole scientific base for knowledge will be shown to be problematic. I will argue that the theory can remain a fruitful option if it is revised in the direction of pluralism.

This article will begin with an analysis of Kornblith's naturalistic epistemology, starting in the Kornblith on Knowledge section where I will present an outline of Kornblith's theory and discuss it in an attempt to elucidate as many relevant aspects as possible. In the An Issue Concerning the Sole Focus on Cognitive Ethology section I will examine a crucial flaw in the theory, regarding Kornblith's choice to solely focus on cognitive ethology as the only science relevant as a base for knowledge. In the Knowledge within Cognitive Neuroscience section the role of knowledge in cognitive neuroscience is presented, as a contrast to Kornblith's focus on cognitive ethology. In the section The Pluralism of Science I will investigate how pluralism, in the context of science, affects Kornblith's theory and present an additional claim that, in my view, saves the theory from the aforementioned flaw while still remaining true to Kornblith's initial stance, followed by a suggested revision of the theory based on this claim in the Revising Kornblith's Theory section. In the Conclusion section a short summary is offered.

Kornblith on Knowledge

Kornblith (2002) argues that many traditional epistemological theories are misconceived. The base for this argument is encapsulated in the following claim:

- (1): '[… T]he subject matter of epistemology is knowledge itself, not our concept of knowledge.' (Kornblith 2002, p. 1)

Since the traditional epistemological focus often is on our intuitions about, or concepts of, different phenomena rather than on the phenomena themselves, most theories can, according to Kornblith, be seen as changing the subject altogether. An investigation into what people *think* about a phenomenon, rather than into the phenomenon itself, might be an interesting, yet distinct, task in its own right (Kornblith 2002, pp. 1–4, 163).

Kornblith's claim points out the possible discrepancy between how the world *is*, and how the world is *believed* to be. It is his view that the world governs truth and falsehood regarding what knowledge is, rather than any intuitions a subject may or may not harbor. If there is a phenomenon of knowledge, then our considerations about it are largely irrelevant for an investigation of the phenomenon. Kornblith *does* acknowledge a role for intuitions, yet views them as inferior to theoretical understanding. Intuitions, often stemming from background knowledge or folk beliefs, *can* be useful in the beginning of a philosophical or scientific investigation, for example by highlighting particularly salient cases, but only until there is better theoretical understanding available, in which case intuitions should give way for empirical investigation.[1]

Where traditional philosophical discussions often focus on intuitions regarding imaginary problems, paradoxes and counterfactual situations, Kornblith's theory affords them merely a preliminary role:

> Intuitions must be taken seriously in the absence of substantial theoretical understanding, but once such theoretical understanding begins to take shape, prior intuitive judgments carry little weight unless they have been endorsed by the progress of theory. The greater one's theoretical understanding, the less weight one may assign untutored judgment. […] Thus, appeal to intuition early on in philosophical investigations should give way to more straightforwardly empirical investigations of external phenomena. (Kornblith 2002, pp. 14–15)

I interpret Kornblith's position on this matter as possible to summarize into a second claim:

- (2): Theoretical understanding trumps intuitive judgment, so intuitions should give way to theoretical understanding based on empirical investigations of external phenomena.

Given this initial discussion, a promising epistemological approach is thus to explore *actual* cases of knowledge, or other relevant phenomena, using the best theoretical understanding available, rather than to investigate what someone happens to find intuitively plausible in a hypothetical situation, at least this is Kornblith's view on the matter.

Distinct Epistemological Questions and Naturalism

According to Kornblith, epistemology should be closely connected to science. However, importantly, epistemology is an autonomous discipline vis-à-vis science since epistemological questions, given their often normative status, frequently differ from scientific questions. So, the questions epistemologists pose ought to be considered legitimate and proper objects of investigation, rather than discarded for being non-

[1] Siegel (2006) criticizes Kornblith for the role he ascribes to intuitions, which he views as question begging. I will grant that Siegel has a point here, although I think that Kornblith's discussion can be seen as offering enough material to answer it. In my view, the issue Siegel raises hinges on whether or not one accepts the stance Kornblith promotes in his third claim. As I interpret Kornblith's argument, he is aware that his line of reasoning demands an acceptance of naturalism. Made evident in his argument and possible to see in formulations such as: 'From *a naturalistic perspective*, there are substantial advantages to looking outward at the phenomena under investigation rather than inward at our intuitions about them.' (Kornblith 2002, p. 16, my italics). What Kornblith wants to do, as I understand him, is not to convince someone who is a firm non-naturalist that he or she *has to* accept that theoretical understanding trumps everyday intuitions, but rather give a plausible explanation of what role intuitions (can) fill in a naturalistic theory. So the question Kornblith discusses is that *given* a naturalistic stance, what role can intuitions play? The question begging that Siegel accuses Kornblith of seems to stem from an interpretation of Kornblith's intentions that is not entirely correct or charitable.

scientific. This means that while Kornblith accepts a rather traditional ontological naturalism, where physical reality is seen as containing nothing "supernatural", he does have a characteristic interpretation of how methodological naturalism should be construed (see, e.g., Papineau 2015; Rysiew 2016). Kornblith is, in my view, best described as endorsing a form of *cooperative* naturalism where epistemologists are allowed to investigate all questions they deem relevant, but need to take scientific findings into account whenever there is theoretical understanding available (Rysiew 2016). So, epistemologists should work with results from science, and also within the boundaries set up by science. The situation can be compared to how, for example, chemistry is constrained by physics, or biology by chemistry (Kornblith 2002, pp. 26–27). This means that Kornblith, given his insistence to ground his theory in science, endorses a form of naturalistic epistemological stance[2]:

- (3): Philosophical investigations ought to adopt a cooperative naturalistic stance.

Although (3) is my interpretation of Kornblith's theory, not something that is openly stated in his text, I believe that the claim is close to Kornblith's view. It is this stance that motivates his approach to philosophy and epistemology.[3] A similar interpretation of Kornblith's theory can be found in Goldman (2005):

> Hilary Kornblith's *Knowledge and Its Place in Nature* has many interesting things to say about what knowledge is and isn't, but its core theses concern meta-epistemology, more broadly, meta-philosophy. Naturalistic epistemology is fundamentally a methodological thesis; it takes a stance on how epistemology should be conducted. Specifically, it holds that epistemology is or should be, in whole or part, an empirical rather than an a priori affair. Kornblith embraces the stronger variant, which says that the subject should be wholly empirical, and this idea is extended to philosophy in general. The book consists of Kornblith's distinctive rationale for this methodological thesis, coupled with many lines of response to naturalism's critics. […T]he core of the book is his detailed program for naturalistic epistemology (and philosophy)[...]. (Goldman 2005, p. 403)

It should be noted that this is not to imply that Kornblith thinks that epistemology, or philosophy, should be *taken over* by science, which would be a *replacement* naturalistic stance, made famous by Quine (1969). According to most interpretations of Quine's classic essay, epistemology is subsumed under cognitive psychology (Quine 1969, p. 82). Since Kornblith's naturalism differs from Quine's, his theory does not face the difficulties that for example Kim (1988) raises for Quine's theory, i.e., that Quine is changing the subject to a focus on causal, rather than justificational, relations (see also Rysiew 2016, section 3.1). Even though Kornblith also has a focus on causal relations, he acknowledges the normative and distinct questions epistemology raise (Kornblith 2002, p. 138).

[2] Kornblith does not elaborate on his version of naturalism, but rather takes it for granted.

[3] To put Kornblith's ideas in context and perspective it might be illuminating to briefly mention that some more or less similar ideas, can be found in for example Maddy (2007) and van Fraassen (2002), who highlight that philosophy should adopt a scientific *attitude* – a stance. However, both Maddy's and van Fraassen's theories differ from Kornblith's on crucial points.

Kornblith also opposes *substantial* naturalism – the view that the questions epistemologists pose should be re-formulated in strictly scientific terminology – and instead sees epistemological questions as legitimate, non-reductive and in need of answers in their own right (Kornblith 2002, pp. 26–27, 171–172; see also Rysiew 2016).

Knowledge as a Natural Phenomenon and Cognitive Ethology

Kornblith argues that to motivate an investigation into any phenomenon, that phenomenon must have a theoretical unity to it. It must be possible to distinguish it from other phenomena. Kornblith argues that knowledge is such a phenomenon:

> There is a robust phenomenon of human knowledge, and a presupposition of the field of epistemology is that cases of knowledge have a good deal of theoretical unity to them; they are not merely some gerrymandered kind, united by nothing more than our willingness to regard them as a kind. […] Now one of the jobs of epistemology, as I see it, is to come to an understanding of this natural phenomenon, human knowledge. (Kornblith 2002, p. 10)

I will extract two claims from the above quote:

- (4): Human knowledge is a natural phenomenon.
- (5): The natural phenomenon of human knowledge has a good deal of theoretical unity.

Kornblith points out that the phenomenon knowledge is, in fact, empirically investigated in science:

> One of the more fruitful areas of such research is cognitive ethology. There is a large literature on animal cognition, and workers in this field typically speak of animals knowing a great many things. They see animal knowledge as a legitimate object of study, a phenomenon with a good deal of theoretical integrity to it. Knowledge, as it is portrayed in this literature, does causal and explanatory work. (Kornblith 2002, pp. 28–29)

I interpret Kornblith's view regarding that cognitive ethology uses knowledge as a causal and explanatory category as an essential claim for his theory:

- (6): Knowledge plays a causal and explanatory role within one of our more fruitful current theories – cognitive ethology.

It now becomes important for Kornblith to show that human knowledge is rightly treated as a form of animal knowledge rather than as separated in kind, since Kornblith sees and uses cognitive ethology as *the* science to investigate both:

> […] I will also argue that human knowledge is not different in kind from the knowledge to be found in the rest of the animal world. Indeed, I will argue that the kind of knowledge that philosophers have talked about all along just is the

> kind of knowledge that cognitive ethologists are currently studying. (Kornblith 2002, pp. 29–30)

This can arguably be summarized into a seventh claim:

- (7): The kind of knowledge that is used in cognitive ethology is also applicable to humans.

It should, however, be noted that there is an ongoing debate regarding anthropomorphism and whether human cognition should be viewed as different in kind or in degree compared to other animals – something Kornblith acknowledges and discusses (Kornblith 2002, pp. 43–48). Kornblith argues that human knowledge should be seen as a form of animal knowledge, at most differing in degree. To motivate his view Kornblith discusses how intentional terminology is widely used in cognitive ethology literature and research, and that it is even necessary to capture some aspects of animal behavior. Intentionality is hence necessary to understand animal behavior according to Kornblith, since descriptions of animal behavior without intentionality merely become descriptions of bodily motions (Kornblith 2002, p. 33). Furthermore, animals seem to need some form of understanding and representation to function in their environment:

> The environment places certain informational demands on an animal. If it is to satisfy its biologically given needs, it will need to recognize certain features of its environment and the evolutionary process must thereby assure that an animal has the cognitive capacities that allow it to deal effectively with that environment. What this requires is the ability to represent information. (Kornblith 2002, p. 37)

The situation described in the above quote makes it possible to attribute mental representations and beliefs to animals as well as humans since it is necessary to make reference to both beliefs and desires to predict both human and animal behavior (Kornblith 2002, p. 42). These aspects can only be fully captured by the intentional terminology used in cognitive ethology:

> There are commonalities among animals that can be captured at the level of talk of belief but cannot be captured in any lower-level vocabulary. […] So when we look at a bit of animal behavior, one question we need to ask is whether its explanation requires talk of informational content, or whether some lower-level explanation, whether chemical or otherwise, will do. (Kornblith 2002, p. 41)

Kornblith gives examples of cognitive ethologists who do ascribe intentionality to animals, and indeed some cognitive ethologists do view human and animal knowledge as similar in kind in Kornblith's sense. However, arguments *against* Kornblith's claim are more plausible than Kornblith is willing to acknowledge. The current state of research suggests that neither view – that human and animal knowledge are relevantly similar or dissimilar – can be ruled out (see, e.g., Klopfer 2005, pp. 204–205). Some issues might ultimately only be possible to settle after a strict definition of key terms, although just how these should be defined might be a matter of theoretical preference and only pushing the problem one step back. If one adopts Kornblith's view, humans are animals among others, and knowledge is a natural phenomenon that humans share with other animals, in which case the differences between human and (other) animal cognitive abilities are just a matter of degree.

Nonetheless, many experiments reach conclusions strengthening the view of human uniqueness, as discussed by Shettleworth (2013, pp. 23–25, 85–88; see also Klopfer 2005, pp. 204–205), among others. Both Shettleworth and Klopfer point out that since many animals have cognitive and sensory abilities that differ a great deal from humans, it might be a mistake to draw too far-reaching conclusions about their similarities (Shettleworth 2013, p. 18).

Wynne (2007) does however point out that most modern ethologists are aware of the risk of anthropomorphism and take this into account in their investigations. Kornblith argues that as long as the fruitfulness of his view trumps other concerns, such as a fear of anthropomorphism, it can be seen as the right approach. Wynne, in the end, is skeptical and fears that anthropomorphism leads to folk-psychological influences that have no scientific relevance (Wynne 2007, p. 134).

According to Kornblith it is *possible* to make a distinction between animal knowledge and human knowledge, since many demarcations are theoretically possible, but it would not mark any significant difference (Kornblith 2002, p. 73).[4] Further aspects of animal and human knowledge can be made evident by examining how self-conscious *reflection* is generally thought to be a central aspect of knowledge – especially human knowledge (Kornblith 2002, p. 103). This theme is elaborated on in Kornblith (2012) in which a more thorough discussion of the topic is carried out. An important point that is highlighted is that introspective justification is often lacking and to a large extent is unreliable, which makes it problematic to let it play any major role in our view on the nature of knowledge. Rather than having a transparent mind, we largely rely on processes beyond our self-conscious, or introspective, grasp. Since many theories of knowledge mark introspection or reflection, in some form, as necessary for – or at least a virtue of – knowledge, this seems to imply that either two forms of knowledge will be needed to meet the different demands, or that different forms of justification need to be accepted to cover all perspectives of the phenomenon of knowledge. Kornblith ultimately argues that introspective reflection and differences in cognitive capacities are non-successful in demarcating human from animal knowledge.

Knowledge as Natural Kind

Natural kinds are, according to Kornblith, to be seen as homeostatically clustered properties, forming a stable unity or a 'well-behaved category' (Kornblith 2002, pp.

[4] Both Kusch (2005) and Bermúdez (2006) question Kornblith's argument against a division between human and animal knowledge, since they claim that even unreflective knowledge – in *humans* – have aspects of logical reasoning built into it. This should, according to Kusch and Bermúdez, be seen as a genuine difference, which Kornblith downplays or ignores. I will regard it to ultimately be an open issue, in that there are arguments both for and against a division. So both interpretations of cognitive ethology and the usage of knowledge regarding animals and humans are reasonable, and the issue is in itself hence not enough to pose any real problem for Kornblith's theory.

61–62). The natural phenomenon knowledge, as instantiated in specific humans or animals, is the locus of such a homeostatic cluster of properties:

> I want to claim that knowledge is, in fact, a natural kind. […] I take natural kinds to be homeostatically clustered properties, properties that are mutually supporting and reinforcing in the face of external change. […] The knowledge that members of a species embody is the locus of a homeostatic cluster of properties; true beliefs that are reliably produced, that are instrumental in the production of behavior successful in meeting biological needs and thereby implicated in the Darwinian explanation of the selective retention of traits. (Kornblith 2002, pp. 61–62)

From this I condense the following claim:

- (8): Knowledge is a natural kind.

Bird and Tobin (2012) describe a natural kind as '[…] a grouping or ordering that does not depend on humans.' (Bird and Tobin 2012), so natural kinds should hence be seen as real groupings in nature, independent of what anybody *thinks* about them. And if one is a scientific realist, as Kornblith is, an investigation using the categories provided by science is the best method there is for understanding what constitutes a natural kind. This is similar to how Kornblith reason concerning the irrelevance of intuitions, and stem from a similar approach, focusing on a phenomenon in nature rather than on people's impressions or intuitions of that phenomenon. So even though a specific scientific theory might be erroneous, there is a fact of the matter concerning the phenomenon. Some traditional examples, often used to show specific natural kinds, are water or H_2O in chemistry and species in biology.

However, Bird and Tobin (2012) mentions that it is somewhat controversial to, for example, speak of natural kinds in biology concerning species – something traditionally thought unproblematic – and that it might be even more so in the social sciences, given that the particulars tend to be more dynamic. Just as regarding anthropomorphism, there is not one particular view that is fully embraced by the scientific community regarding natural kinds. Kornblith could once more be seen to downplay a debate that has far from reached a conclusive scientific consensus and instead presents his view concerning natural kinds, and knowledge as a natural kind, as less complicated than it is.[5] There might be many acceptable ways to classify the world, and the same phenomenon in it, into kinds and perhaps still to regard them as natural kinds.

[5] Bermúdez (2006) points out cases where cognitive ethologists disagree with Kornblith's main tenets and about the possibility of using knowledge as a natural kind. I do not question Bermúdez in his argumentation and examples regarding other interpretations of how cognitive ethology should be viewed. But as concerning the previous point of anthropomorphism there is no general interpretation of the results from cognitive ethology that is totally conclusive and accepted by the majority of research, so I do not think that this is enough to pose a real threat to Kornblith's theory.

Knowledge Requiring Reliably Produced True Belief (RTB)

According to Kornblith we should look to cognitive ethology for an understanding of knowledge, and cognitive ethology tells us that:

> Knowledge explains the possibility of successful behavior in an environment, which in turn explains fitness. [… W]e must appeal to a capacity to recognize features of the environment, and thus the true beliefs that [… someone] acquire will be the product of a stable capacity for the production of true beliefs. The resulting true beliefs are not merely accidentally true; they are produced by a cognitive capacity that is attuned to its environment. In a word, the beliefs are reliably produced. The concept of knowledge which is of interest here thus requires reliably produced true belief. (Kornblith 2002, pp. 57–58)

Kornblith's interpretation of cognitive ethology leads him to the following claim:

- (9): 'Knowledge is a robust category in the ethology literature; it is more than belief, and more than true belief. It requires reliably produced true belief.' (Kornblith 2002, p. 69)

Even though I consider the following claim in need of further discussion, which I will present below, Kornblith explicitly states:

- (10): 'The conception of knowledge that we derived from cognitive ethology literature, a reliabilist conception of knowledge, gives us the *only* viable account of what knowledge is.' (Kornblith 2002, p. 135, my italics)

Tying together all previously mentioned claims, (1)–(10), I argue that we arrive at the following conclusion:

– (*i*): Reliabilist knowledge, requiring *RTB*, is the only viable account of what knowledge is.

Above I have tried to present and discuss Kornblith's naturalistic epistemological theory as a framework consisting of ten claims and a conclusion regarding what knowledge is. Claim (3) does stand out from the other claims in that it is normative. As previously mentioned, my interpretation of Kornblith's theory is that it promotes a cooperative naturalistic stance about how epistemology – and philosophy – ought to be conducted, which affects how we ought to view knowledge.

An Issue Concerning the Sole Focus on Cognitive Ethology

Kusch (2005) raises an issue that is genuinely problematic for Kornblith's theory. This issue, in my view, is so serious that Kornblith's theory in its present state should be abandoned. That said, I find that Kornblith's theory has so many fruitful aspects and strengths that it is worthwhile to consider possible revisions. In short, Kusch points out that it seems questionable to let cognitive ethology give us the only viable account of what knowledge is, when other sciences see knowledge in other ways:

> Kornblith rightly insists that the best way to find out about knowledge is to turn to scientific enquiry. He writes: 'Where should we turn, and how should we proceed, if we are to investigate the phenomenon of knowledge itself? … One of the most fruitful areas of such research is cognitive ethology....' (28). Unfortunately, it turns out that this is the *only area* of 'such research' to which Kornblith pays attention. A critical reader cannot but wonder why cognitive ethology receives this special position. […] Which account of knowledge should we favour: the account offered by cognitive ethology or the account proposed by the sociology of scientific knowledge? I see no reason to prefer one over the other. (Kusch 2005, pp. 414–415)[6]

The sociology of scientific knowledge, upon which Kusch's criticism focuses, is a scientific field that investigates science as a social phenomenon. It is closely related to both sociology and the sociology of knowledge and emphasizes social factors and the cultural context surrounding a research paradigm, presented and discussed by Shapin (1995) and others. Kusch argues that knowledge, from the perspective of the sociology of scientific knowledge, might be viewed as a *social kind*. Since Kornblith's theory is a version of naturalistic realism and the sociology of scientific knowledge relates more readily with anti-realism, the two theories can be seen as endorsing two quite different stances.

Kornblith (2005, see also 2006) presents a reply to Kusch, discussing why cognitive ethology's take on knowledge is preferable to that of the sociology of scientific knowledge, also addressing other criticisms raised by Kusch. But regardless of whether Kornblith's rebuttal of the sociology of scientific knowledge is accepted or not, he sidesteps the more overarching issue regarding why sciences other than cognitive ethology should be disallowed. Even if Kornblith's stance is adopted, and cognitive ethology is seen as preferable to the sociology of scientific knowledge, the step from seeing cognitive ethology as *one* possible science of interest to it being the *only* one is not properly motivated – in the original text or in his reply to Kusch. Kornblith does not, for example, investigate how different sciences closer to his naturalistic realistic stance invoke knowledge. In his argumentation regarding human and animal knowledge, discussed in the Knowledge as a Natural Phenomenon and Cognitive Ethology section above, Kornblith briefly mentions how lower-level explanations of intentional phenomena risks missing central aspects that higher-level explanations are better suited to deal with, by abstracting away from physical details (Kornblith 2002, pp. 39–41; see also Kornblith 1993, pp. 54–57). An anti-reductionist position regarding higher-level theories about natual phenomena such as knowledge, might allow us to abstract away from (some) physical micro-details in certain contexts, but this would arguably not by itself make all lower-level sciences illegitimate. To let philosophy – or epistemology – be the arbiter of which sciences we should take seriously or not seems to be at odds with the cooperative naturalistic stance, and is something Kornblith explicitly warns

[6] Kornblith actually writes that cognitive ethology is 'One of the *more* fruitful areas […]' (Kornblith 2002, p. 28, my italics).

against (Kornblith 2002, p. 32).[7] Nothing in Kornblith's line of reasoning indicates why we should ignore or invalidate all sciences other than cognitive ethology. What can be assessed is that (10), the claim that cognitive ethology gives us the only viable account of knowledge, is not convincingly motivated.

Knowledge within Cognitive Neuroscience

In this section I will focus on another scientific field in which knowledge plays an essential role, apart from cognitive ethology and the sociology of scientific knowledge, namely cognitive neuroscience. I will show that knowledge is used as a category that plays a causal and explanatory role within this field as well, which lies closer to cognitive ethology than the sociology of scientific knowledge. The significance of this discussion is that the constraints that Kornblith puts on knowledge in (10) become even more questionable: the issue concerning the sole focus on cognitive ethology remains even if knowledge is seen as a natural rather than a social kind.

According to cognitive neuroscience – a diverse field studying the biological foundations of cognitive processes – people are considered to get information from their senses, whereas the information is comprehended only after a complex combination of processes that leads to perceptions (for a comprehensive overview see, e.g., Bickle 2009). This means that we cannot directly understand information that reaches our sense organs, which in itself is not comprehensible to us. Rather, we need to process the information that reaches us before the information becomes meaningful perceptions from which we can reason and act (Gazzaniga et al. 2002; see also Friston 2009, 2010).

Long-term memory (LTM) is conventionally seen as the most relevant function(s) of the brain for the analysis of knowledge. LTM is commonly divided into the nested categories procedural memory, semantic memory and episodic memory (see, e.g., Tulving 1985), and is thought to be able to handle a, practically speaking, infinite amount of information. LTM is grouped into two main categories: non-declarative (or implicit, non-accessible) memory, and declarative (or explicit, accessible) memory. Non-declarative procedural memory, beyond our conscious reach, handles our ability to perform actions, whereas consciously aware declarative semantic memory handles categorizations and concepts, and episodic memory handles remembered events and facts. Knowledge is in the traditional philosophical debate commonly divided into procedural knowledge and propositional knowledge, which in the cognitive neuroscientific terminology maps to procedural memory and episodic memory respectively. The examples below will however focus on conceptual knowledge, which maps to semantic memory.

To show that knowledge plays a causal and explanatory role in cognitive neuroscience, I will cite what I consider to be representative passages from cognitive neuroscientific texts. Pursuing clarity, I will only focus on semantic memory and conceptual knowledge. However, a similar presentation could easily be given concerning procedural memory and procedural knowledge or concerning episodic memory and propositional knowledge. More detailed arguments and discussions concerning different

[7] I will reconnect to this point below.

specific neuroscientific theories can be found in for example Churchland (1986), Bennett and Hacker (2003) and Bennett et al. (2007).

In the words of Gazzaniga et al., semantic memories are described as:

> World knowledge, object knowledge, language knowledge, conceptual priming. (Gazzaniga et al. 2002, p. 314)

Connecting semantic memory with knowledge, Ward (2010) writes that:

> Semantic memory is conceptually based knowledge about the world, including knowledge of people, places, the meaning of objects and words. It is culturally shared knowledge. By contrast, episodic memory refers to memory of specific events in one's own life. The memories are specific in time and place. For example, knowing that Paris is the capital of France is semantic memory, but remembering a visit to Paris or remembering being taught this fact is episodic memory. (Ward 2010, p. 186)

Patterson et al. (2007) give the following description of semantic memory and knowledge:

> Semantic memory (also called conceptual knowledge) is the aspect of human memory that corresponds to general knowledge of objects, word meanings, facts and people, without connection to any particular time or place. (Patterson et al. 2007, p. 976)

Binder and Desai (2011) give this account of semantic memory:

> […] semantic memory is one of our most defining human traits, encompassing all the declarative knowledge we acquire about the world. A short list of examples includes the names and physical attributes of all objects, the origin and history of objects, the names and attributes of actions, all abstract concepts and their names, knowledge of how people behave and why, opinions and beliefs, knowledge of historical events, knowledge of causes and effects, associations between concepts, categories and their bases, and on and on. […] All of human culture, including science, literature, social institutions, religion, and art, is constructed from conceptual knowledge. We do not reason, plan the future or remember the past without conceptual content – all of these activities depend on activation of concepts stored in semantic memory. (Binder and Desai 2011, p. 527)

Yee et al. (2014) describe their view of semantic memory and knowledge:

> How do we know what we know about the world? For instance, how do we know that a cup must be concave, or that a lemon is normally yellow and sour? Psychologists and cognitive neuroscientists use the term *semantic memory* to refer to this kind of world knowledge. […] Today, most psychologists use the term *semantic memory* […]—to refer to all kinds of general world knowledge, whether it be about words or concepts, facts or beliefs. What these types of world

> knowledge have in common is that they are made up of knowledge that is independent of specific experiences; instead, it is general information or knowledge that can be retrieved without reference to the circumstances in which it was originally acquired. (Yee et al. 2014, p. 353)

As can be seen from this quote, and the next, it is possible to interpret Yee et al. as using semantic knowledge and semantic memory interchangeably. Furthermore, knowledge is used as a category to investigate the causal underpinnings of the memory system:

> Thus, the evidence suggests that semantic knowledge can be acquired independently of the episodic memory system. However, semantic knowledge in these amnesic patients is not normal (e.g., it is acquired very slowly and laboriously). It is therefore possible that the acquisition of semantic memory normally depends on the episodic system, but other points of entry can be used (albeit less efficiently) when the episodic system is damaged. Alternatively, these patients may have enough remaining episodic memory to allow the acquisition of semantic knowledge (Squire and Zola, 1998). (Yee et al. 2014, p. 354)

So, knowledge does indeed play a causal and explanatory role in cognitive neuroscience – as it does in cognitive ethology. But, cognitive ethology has an *ultimate* focus on *why* a behavior occurs and on what animals should do, whereas cognitive neuroscience has a *proximate* focus on *how* animals do what they do (Scott-Phillips et al. 2011; Martin and Bateson 2007; Tinbergen 1963). This divergence leads to a situation where knowledge as understood in cognitive ethology requires reliably produced true belief (9), whereas knowledge as understood in cognitive neuroscience *is* LTM.

Elaborating on this divergence, and speaking against the compatibility of the two perspectives, the *un*reliability of human cognition and memory can be pointed out. For example, Tversky and Kahneman (1971, 1974) show how people tend to consistently make errors in their representations and inferences in some situations. These, and similar findings (see, e.g., Nisbett and Borgida 1975; Ross et al. 1975), indicates that LTM does in fact not readily provide reliable true belief, and that knowledge hence cannot be seen as requiring this, since LTM *is* knowledge, from a cognitive neuroscientific perspective. LTM might sometimes and under certain circumstances provide reliable true belief, but at other times, and under other circumstances, this might not be the case.

An argument supporting Kornblith's position indicating compatibility between reliable true belief and LTM might instead emphasize how the above point only applies in contrived situations and that animals (including humans) have an evolutionarily grounded tendency to come out right in their generalizations and predictions:

> Knowledge may never be absolute and certain, but it is always true enough to be workable. (Plotkin 1993, p. 121)

However, even if the two sciences are seen as compatible, my point is that the two perspectives *do* diverge in important ways and that it is untenable to only allow the ultimate perspective as a base for giving us a viable account of what knowledge is.

From a naturalistic perspective, as pointed out in the section An Issue Concerning the Sole Focus on Cognitive Ethology, it is not the role of philosophy to pit different sciences against each other or to judge which sciences we should dismiss or follow, making Kornblith's claim (10) insupportable.[8]

The Pluralism of Science

There are actually a number of interconnecting sciences inquiring into animal cognition, and hence at least potentially into 'knowledge', for example, cognitive neuroscience, developmental psychology, neurobiology, cognitive psychology, cognitive ethology, behavioral ecology, evolutionary psychology, evolutionary biology and cognitive zoology. However, for the purposes of the present argumentation it suffices to note that cognitive neuroscience belongs to this group.

Dupré (1993) argues that science cannot be seen as a unified project, since the world consists of such overwhelming pluralistic diversity. Any phenomenon is, according to Durpé, possible to reduce to multiple different natural kinds, depending on the context and goal that is seen as relevant (Dupré 1993, pp. 1–5). What is to be considered a natural kind therefore depends on context, which in turn hinges on the goals of an investigator. Focusing on Kornblith's theory, it can only be said to identify knowledge as a natural kind given a particular context and goal. From this perspective, Kornblith is in effect unreasonably excluding the possibility that other sciences could investigate the phenomenon from their particular context and with their goals.

A similar, albeit distinctly different, position is offered by Horst (see, e.g., Horst 2011, Horst 2016), who points out that all scientific models have some degree of idealization and abstraction built into them. The diversity and disunity Dupré ascribes to the world could thus instead be interpreted as a result of disunities in how we *model* the world (Horst 2011, p. 69):

> [… T]he mind employs a plurality of mental models, […] each idealized in form, and consequently […] scientific models of any of these mental models must be viewed as partial and idealized. (Horst 2011, p. 254)

Horst offers an interesting framework for scientific theories and models, which he calls 'cognitive pluralism':

> Within a Cognitive Pluralist framework, however, we can see these as variations on a theme rather than as essential differences. All models are plural, partial, idealized, and cast in some particular representational system. Scientific models are particularly regimented and formally exact. And within the class of scientific models we find different types of idealization conditions that result in closer or

[8] An argument for the priority of cognitive ethology over cognitive neuroscience might be found in the thesis of multiple realizability, where cognitive ethology can be interpreted as better equipped to explain what knowledge is given its more functionalistic ultimate perspective. However, if the differences between humans and other animals are made salient, the same thesis can just as well be used against Kornblith's earlier merging of human and animal knowledge, and instead be interpreted as pointing out the importance of species-specific differences.

> more distant relationships between models and the real-world behavior that they are invoked to explain. (Horst 2011, p. 261)

Just how we model a natural phenomenon, such as knowledge, will thus hinge on which science we use, without necessarily saying anything about the underlying properties – diverging models are possible of the same natural kind. In other words we can investigate and try to "triangulate" the same natural kind – the homeostatically clustered properties forming a well-behaved category – by looking at it through different "lenses", which all might skew our view in idiosyncratic ways resulting in diverging accounts of the same phenomenon (see, e.g., Horst 2016, p. 83).

In fact, support for a pluralistic way of thinking about natural kinds can be found in Kornblith's own work:

> Not just any scheme of classification corresponds to the real kinds in nature. It is just that the structure of the real kinds may not be as simple or as neat as has been dreamt of in many philosophies. The homeostatic cluster account thus suggests a rich overlapping structure of kinds in nature, with the various sciences picking out families of kinds which are interrelated. (Kornblith 1993, p. 52)

The step from granting that different sciences pick out families of kinds that are interrelated to granting that this is so in the case of knowledge is very short indeed. Scientific pluralism and theoretical unity can on such an account, in my view, be seen as compatible. Knowledge can hence be interpreted as to consist of a slightly more inclusive overlapping and interrelated structure than is ordinarily assumed. The various sciences' accounts of the natural kind will accordingly be affected by their particular "lens" and be more or less commensurable (Horst 2016, pp. 7, 222–226).[9]

As previously mentioned, the different sciences relevant in regards to animal cognition focus on partly different aspects, or points of view; for example, cognitive neuroscience, developmental psychology, neurobiology and cognitive psychology have a proximate focus on *how* animals do what they do, whereas cognitive ethology, behavioral ecology and evolutionary psychology have an ultimate focus on *why* a behavior occurs and what animals should do (Scott-Phillips et al. 2011). Scott-Philips et al. points out the importance of clearly stating the framework from within which one works, and the possibility of investigating the same phenomenon from multiple points of view.

The above ideas regarding pluralism and the importance of different points of view can be given a firmer standing with the help of the concept of 'levels'. The world can be investigated at different levels, for example from the perspective of: physics, chemistry, cellular biology, functional biology, psychology, sociology, and so on. To illustrate the different "middle-range" levels, and how they affect our view of knowledge, at least four different sciences come readily to mind: cognitive neuroscience, cognitive psychology, the sociology of scientific knowledge and cognitive ethology. Of these four levels Kornblith favors the latter, Kusch favors the third and I have discussed the first above. But in all of the above-mentioned scientific fields it can be argued that knowledge plays an important role – just as it does in cognitive ethology – and is treated as a phenomenon with theoretical unity. The key issue here is, in my view,

[9] If this is not taken into account theoreticians risk talking past each other.

whether one favors a more traditional top-down approach focusing on "higher" functions, in which case cognitive ethology is a natural choice of science to focus on. If one, on the other hand, favors a bottom-up approach focusing on how the "lower" levels affect the higher ones, cognitive neuroscience is an interesting candidate. LTM could then be seen as knowledge on a cognitive neuroscientific level of explanation, and as the underlying microstructure for knowledge (*RTB*) on the higher cognitive ethological level of explanation.[10]

A more inclusive version of claim (6) thus ought to be introduced, that allows for all relevant sciences to be used in an investigation:

- (6*i*): Knowledge plays a causal and explanatory role within several of our more fruitful current theories.

As the previous discussion has shown, it is relevant to take context and goals, point of view, and level of explanation into account while investigating knowledge. Depending on how one chooses to position oneself concerning these matters, investigations will take different forms and different sciences will be more or less relevant. To enable a pluralistic revision of Kornblith's theory the following claim should thus be added:

- (11): Philosophical investigations ought to adopt a stance accommodating scientific pluralism.

Revising Kornblith's Theory

From the above discussion it should be clear that at least cognitive neuroscience is a legitimate science in which knowledge plays an essential role, and yet its account of what knowledge is diverges from the account found in cognitive ethology, regarding context, goals, focus and level of explanation. Kornblith's claim that cognitive ethology gives us the *only* viable account of knowledge is thus not plausible. Kornblith's theory needs to be revised, along the lines already proposed, in order to save the theory from the issue concerning the sole focus on cognitive ethology.

To be concrete, Kornblith needs to retract claim (10) as well as conclusion (*i*).[11] What then follows is that reliabilist knowledge, requiring *RTB*, is *one* viable account of what knowledge is. This is a plausible conclusion given that one's focus is on cognitive ethology. However, if we replace (6) by (6*i*), as previously hinted, and add claim (11) while removing claims (7) and (9), what we get is the kind of pluralism which our argument has led us to:

- (1): The subject matter of epistemology is knowledge itself, not our concept of knowledge.

[10] A contrasting opinion and discussion can be found in for example Horvath (2016, pp. 175–176).

[11] Bermúdez mentions similar concerns, but sees the situation facing Kornblith's theory as risking it being dubbed folk psychology rather than focusing on the possibility of an inclusive pluralism (Bermúdez 2006, p. 304).

- (2): Theoretical understanding trumps intuitive judgment, so intuitions should give way to theoretical understanding based on empirical investigations of external phenomena.
- (3): Philosophical investigations ought to adopt a cooperative naturalistic stance.
- (4): Human knowledge is a natural phenomenon.
- (5): The natural phenomenon of human knowledge has a good deal of theoretical unity.
- (6*i*): Knowledge plays a causal and explanatory role within several of our more fruitful current theories.
- (8): Knowledge is a natural kind.
- (11): Philosophical investigations ought to adopt a stance accommodating scientific pluralism.

The theory thus outlined retains important insights of Kornblith's theory while, at the same time, saving that theory from the issue concerning the sole focus on cognitive ethology.

Conclusion

I addressed Hilary Kornblith's proposal that knowledge is a natural kind, the identification of which is the unique responsibility of one particular science: cognitive ethology. As Kornblith sees it, the natural kind thus picked out is knowledge as construed by reliabilism. I have argued that knowledge plays a causal and explanatory role within many of our more fruitful current theories, diverging from the reliabilist conception even in disciplines that are closely related to cognitive ethology, focusing on cognitive neuroscience. Rather than discarding the natural kind approach altogether, as some authors have been tempted to do, I proposed that many of Kornblith's insights can in fact be preserved within a framework that is both naturalist and pluralist. In this way Kornblith's naturalistic epistemology, in its revised pluralist form, can remain a promising and fruitful framework for investigating knowledge – indeed as a natural kind.

Acknowledgments I am very grateful to Erik J. Olsson for his guidance, helpful comments and for him being extremely charitable with his knowledge and time. Thanks to Johannes Persson, Shira Singer and my anonymous referees for comments.

References

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Malden, MA: Blackwell Publishers.

Bennett, M., Dennett, D., Hacker, P., & Searle, J. (2007). *Neuroscience and Philosophy: Brain, Mind, and Language*. New York: Columbia University Press.

Bermúdez, J. L. (2006). Knowledge, Naturalism, and Cognitive Ethology: Kornblith's *Knowledge and its Place in Nature*. *Philosophical Studies, 127*, 299–316.

Bickle, J. (ed.) (2009). *The Oxford Handbook of Philosophy and Neuroscience*. Oxford: Oxford University Press.

Binder, J. R., & Desai, R. H. (2011). The Neurobiology of Semantic Memory. *Trends in Cognitive Sciences, 15*(11), 527–536.

Bird, A. and Tobin, E. (2012). 'Natural Kinds', In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition). <http://plato.stanford.edu/archives/win2012/entries/natural-kinds/>.

Bonjour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Chisholm, R. M. (1988). The Indispensability of Internal Justification. *Synthese, 74*, 285–296.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: The MIT Press.

Churchland, P. M., & Churchland, P. S. (1998). *On the Contrary: Critical Essays 1987–1997*. Cambridge, MA: The MIT Press.

Dennett, D. (1996). *The Intentional Stance*. Cambridge, MA: The MIT Press.

Dewey, J. (1938). *Logic: The Theory of Inquiry.* New York: Holt, Rinehart and Winston.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: The MIT Press.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Friston, K. (2009). The Free-Energy Principle: A Rough Guide to the Brain? *Trends in Cognitive Sciences, 13*(7), 293–301.

Friston, K. (2010). The Free-Energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive Neuroscience: The Biology of the Mind*. New York: W. W. Norton & Company.

Goldman, A. I. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. I. (2005). Kornblith's Naturalistic Epistemology. *Philosophy and Phenomenological Research, 71*(2), 403–410.

Horst, S. (2011). *Laws, Mind, and Free Will*. Cambridge, MA: The MIT Press.

Horst, S. (2016). *Cognitive Pluralism*. Cambridge, MA: The MIT Press.

Horvath, J. (2016). Conceptual Analysis and Natural Kinds: The Case of Knowledge. *Synthese, 193*(1), 167–184.

James, W. (1907 [1995]). *Pragmatism*: *A New Name for Some Old Ways of Thinking*. New York: Dover.

Kim, J. (1988). What is "Naturalized Epistemology"? *Philosophical Perspectives, 2*, 381–405.

Klopfer, P. H. (2005). Animal Cognition and the New Anthropomorphism. *International Journal of Comparative Psychology, 18*(3), 202–206.

Kornblith, H. (1993). *Inductive Inference and Its Natural Ground: An Essay in Naturalistic Epistemology.* Cambridge, MA: The MIT Press.

Kornblith, H. (2002). *Knowledge and its Place in Nature.* Oxford: Oxford University Press.

Kornblith, H. (2005). Replies to Alvin Goldman, Martin Kusch and William Talbott. *Philosophy and Phenomenological Research, 71*(2), 427–441.

Kornblith, H. (2006). Reply to Bermúdez and BonJour. *Philosophical Studies, 127*, 337–349.

Kornblith, H. (2012). *On Reflection*. Oxford: Oxford University Press.

Kusch, M. (2005). Beliefs, Kinds and Rules: A Comment on Kornblith's *Knowledge and Its Place in Nature*. *Philosophy and Phenomenological Research, 71*(2), 411–419.

Maddy, P. (2007). *Second Philosophy: A Naturalistic Method*. Oxford: Oxford University Press.

Martin, P., & Bateson, P. (2007). *Measuring Behaviour.* Cambridge: Cambridge University Press.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32*(5), 932–943.

Papineau, D. (2015). 'Naturalism', In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition). <http://plato.stanford.edu/archives/fall2015/entries/naturalism/>

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do You Know what you Know? The Representation of Semantic Knowledge in the Human Brain. *Nature Review Neuroscience, 8*, 976–987.

Peirce, C. S. (1877 [1955]). 'The Fixation of Belief', In J. Buchler (ed.), Philosophical Writings of Peirce (pp. 5–22). New York: Dover Publications, Inc.

Plotkin, H. C. (1993). *Darwin Machines and the Nature of Knowledge*. Cambridge, MA: Harvard University Press.

Quine, W. V. O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays* (pp. 69–90). New York: Columbia University Press.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*, 880–892.

Rysiew, P. (2016). ‘Naturalism in Epistemology’, In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), forthcoming. <http://plato.stanford.edu/archives/sum2016/entries/epistemology-naturalized/>

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary Theory and the Ultimate-Proximate Distinction in the Human Behavioral Sciences. *Perspectives on Psychological Science, 6*(1), 38–47.

Shapin, S. (1995). Here and Everywhere: Sociology of Scientific Knowledge. *Annual Review of Sociology, 21*, 289–321.

Shettleworth, S. J. (2013). *Fundamentals of Comparative Cognition*. Oxford: Oxford University Press.

Siegel, H. (2006). Book Review: Hilary Kornblith, *Knowledge and its Place in Nature*. *Philosophical Review, 115*(2), 246–251.

Tinbergen, N. (1963). On Aims and Methods of Ethology. *Zeitschrift für Tierpsychologie, 20*, 410–433.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1–12.

Tversky, A., & Kahneman, D. (1971). Belief in the Law of Small Numbers. *Psychological Bulletin, 76*(2), 105–110.

Tversky, A., & Kahneman, D. (1974). ‘Judgment under Uncertainty: Heuristics and Biases’, *Science*. *New Series, 185*(4157), 1124–1131.

van Fraassen, B. C. (2002). *The Empirical Stance*. London: Yale University Press.

Ward, J. (2010). *The Student's Guide to Cognitive Neuroscience*. New York: Psychology Press.

Wynne, C. D. L. (2007). What are Animals? Why Anthropomorphism is still not a Scientific Approach to Behavior. *Comparative Cognition and Behavior Reviews, 2*, 125–135.

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2014). Semantic Memory. In K. Ochsner & S. Kosslyn (eds.), *The Oxford Handbook of Cognitive Neuroscience: Volume 1, Core Topics* (pp. 353–374). Oxford: Oxford University Press.

Paper II

P. Gärdenfors, A. Stephens
Induction and knowledge-what
European Journal for Philosophy of Science, 2017 [2018], *8*(3), 471-49

Euro Jnl Phil Sci (2018) 8:471–491
https://doi.org/10.1007/s13194-017-0196-y

ORIGINAL PAPER IN FORMAL EPISTEMOLOGY

Induction and knowledge-what

Peter Gärdenfors[1] · **Andreas Stephens**[1]

Received: 21 April 2017 / Accepted: 23 November 2017 / Published online: 13 December 2017

Abstract Within analytic philosophy, induction has been seen as a problem concerning inferences that have been analysed as relations between sentences. In this article, we argue that induction does not primarily concern relations between sentences, but between *properties* and *categories*. We outline a new approach to induction that is based on two theses. The first thesis is epistemological. We submit that there is not only knowledge-how and knowledge-that, but also *knowledge-what*. Knowledge-what concerns relations between properties and categories and we argue that it cannot be reduced to knowledge-that. We support the partition of knowledge by mapping it onto the long-term memory systems: procedural, semantic and episodic memory. The second thesis is that the role of inductive reasoning is to generate knowledge-what. We use conceptual spaces to model knowledge-what and the relations between properties and categories involved in induction.

Keywords Inductive inferences · Knowledge-what · Semantic memory · Conceptual knowledge · Naturalistic epistemology

1 Introduction

One of the most impressive features of human cognitive processing is our ability to perform *inductive inferences*. We generalise from a very limited number of observations, sometimes with overwhelming confidence. A central problem in philosophy of science concerns how the mechanism of inductive reasoning can be described and motivated.

✉ Peter Gärdenfors
peter.gardenfors@lucs.lu.se

Andreas Stephens
andreas.stephens@gmail.com

[1] Department of Philosophy, Lund University, Lund, Sweden

We do not perform inductive inferences in an arbitrary manner. Peirce notes that there are certain forms of constraints that delimit the vast class of possible inferences. As he puts it:

> Nature is a far vaster and less clearly arranged repertory of facts than a census report; and if men had not come to it with special aptitudes for guessing right, it may well be doubted whether in the ten or twenty thousand years that they may have existed their greatest mind would have attained the amount of knowledge which is actually possessed by the lowest idiot. But, in point of fact, not man merely, but all animals derive by inheritance (presumably by natural selection) two classes of ideas which adapt them to their environment. In the first place, they all have from birth some notions, however crude and concrete, of force, matter, space, and time; and, in the next place, they have some notion of what sort of objects their fellow-beings are, and how they will act on given occasions. (Peirce 1955, pp. 214–5)

Here, Peirce hints at an *evolutionary* explanation of why "the human intellect is peculiarly adapted to the comprehension of the laws and facts of nature" (Peirce 1955, p. 213).

Within analytic philosophy, induction has been seen as a problem concerning inferences that have been analysed as relations between sentences. Inductive inferences were important for the logical positivists, being a cardinal component in their verificationist program (see, e.g., Carnap 1950; Hempel 1965; Rosenberg 2000; Ladyman 2002; Creath 2014; Vickers 2014). However, it soon became apparent that their logical approach resulted in paradoxes. The most well-known are Hempel's (1965) 'paradox of confirmation' and Goodman's (1983) 'new riddle of induction'. If we use logical relations alone to determine which inductions are valid, the fact that all predicates are treated on a par induces *symmetries* which are not preserved by our intuitions concerning which inductive inferences are permissible: 'Raven' in Hempel's paradox is treated on a par with 'non-raven', 'green' in Goodman's with 'grue', etc. What is needed is a non-logical way of distinguishing the predicates that may be used in inductive inferences from those that may not.

In this article, our diagnosis of why the paradoxes have emerged in the traditional treatment is that induction does not primarily concern relations between sentences, but between *properties* and *categories*. We outline a new approach to induction that is based on two theses. The first one is epistemological. We argue that there is not only knowledge-how and knowledge-that, but also *knowledge-what*. Knowledge-what concerns relations between properties and categories and we argue that it cannot be reduced to knowledge-that. We motivate our approach by giving it a naturalistic grounding in cognitive neuroscience.

The second thesis is that the role of induction is to generate knowledge-what. This entails that we find much of the earlier discussion of induction misguided since it has focused on induction as generalisations generating knowledge-that. In this context, it should be noted that there are two meanings of 'generalisation' in the literature. One is logical, relating to the connection between sentences describing individual instances and universal sentences covering the individual sentences. The other, also called 'stimulus generalisation', is psychological and concerns the relations between reactions to a particular stimulus and a class of *similar* stimuli. We argue that human inductive inference is more related to the psychological notion of generalisation.

A central question then is how knowledge-what can be modelled. Here we build on the theory of conceptual spaces proposed by Gärdenfors (1990, 2000, 2014). In this theory, knowledge is organised into domains modelled as spatial structures. Properties are analysed as (convex) regions within such domains and categories as complexes of regions from different domains. There are several dimensional theories of categorisation, but the unique property of the theory of conceptual spaces is its strong reliance on geometric structures.

Before we present our analysis of knowledge-what and its relation to induction, we give, in section 2, a brief account of why induction has been seen as generating knowledge-that, and then outline our own cognitivist and naturalistic stance. We present arguments for dividing knowledge into knowledge-how, knowledge-what and knowledge-that in section 3. In section 4, we map this tripartition of knowledge onto three kinds of long-term memory – procedural, semantic and episodic – thereby connecting our account of knowledge to cognitive neuroscience. Then in section 5 we introduce conceptual spaces as a tool for modelling knowledge-what in the form of relations between categories and properties. Finally, in section 6 we argue that induction concerns methods for generating knowledge-what.

2 Two approaches to induction

In this section, we sketch an account of why inductive inferences have been seen as relations between sentences, and then present our alternative naturalistic approach. We derive the approaches from the underlying views of what constitute knowledge.

2.1 Induction from the perspective of language

Historically, the empiricist turn of the seventeenth century raised an interest in inductive inferences, although it remained uncertain how induction should be justified since it lacked the logical rigor of deduction. Hume (1988) argued that it is impossible to justify inductive inferences, although he acknowledged *habit* as an inevitable part of human reasoning. Other issues included pinpointing which evidence, and what amount, was enough for valid inductive inferences as well as finding methods that could separate good inferences from bad ones (see, e.g., Mill 1843; Vickers 2014).

In the mainstream debate within analytic philosophy, a major distinction has been that between *knowledge-how* and *knowledge-that* (Ryle 1949). It has been a tacit assumption that induction does not concern knowledge-how. As part of the linguistic turn of analytic philosophy, there was a preference for analysing inferences, including induction, as relations between sentences. Hence, it was concluded that if induction is an epistemic process, it must deal with knowledge-that, since knowledge-that is propositional and can be expressed in sentences.

For the logical positivists, the basic objects of study were sentences in some more or less regimented language. Ideally, the language was a version of first-order logic where the atomic predicates represent observational properties. These observational predicates were taken as primitive, unanalysable notions. The main tool used when studying the linguistic expressions was logical analysis. In its purest form, logical positivism allowed only this tool. A consequence of this methodology was that all observational predicates were treated in the same way since there were no logical reasons to differentiate between them. For example, Carnap (1950, sec. 18B) required that the primitive predicates of a language be logically independent of each other.

In this tradition, particular observations are used as evidence for inductive generalisations or predictions (Carnap 1950, 1971; Hempel 1965). When connections are found within the registered observations, inductive generalisations can be made, which then can be confirmed by additional observations. So if observations of objects O_1, O_2, O_3 … all are *C*, the generalisation that *all O* are *C* can be made. The inference thus concerns a relation between individual and universal sentences. The evidence from the premises gives stronger or weaker support for the conclusion. Since this inductive process does not have the same logical rigor as a deductive process, the methodology of induction requires that supporting evidence preferably should come in large numbers, come from several different contexts and have no negative cases (Hempel 1965).

One point that has been downplayed in the debate, however, is that not all universal sentences can function as conclusions in inductive inferences. Ever since Aristotle's classic "All men are mortal", inductive inferences only involve universal sentences that are *generics*, that is, express relations between categories and properties. A non-generic universal sentence such as "All persons in this room are Swedish" would not be acceptable as an inductive inference, even when perfectly supported by the given evidence. This is so since such 'accidental generalisations' do not support counterfactuals of the form "if a person came into the room he or she would be a Swede". So, even though the logical form of a law-like sentence is the same as that of an accidental universal sentence, we point to the connection between law-like sentences and generics. In the literature there has been attempts to distinguish 'law-like' (nomologic, nomothetic) generalisations from 'accidental' generalisations (see, e.g., Goodman 1983; Hempel 1965), and early steps to break the logical emphasis were taken by for example Dretske (1977), Tooley (1977) and Armstrong (1978, 1983) who focused on laws as relations of non-logical *necessitation* between universals (see also Carroll 2016).

2.2 Induction from a naturalistic perspective

Our alternative to the traditional propositional or sentential approach is *cognitivist* and *naturalistic*. We thus highlight that inductive inferences are possible since the world has moulded our cognitive faculties through evolution (Quine 1969b; Lorenz 1977; Gärdenfors 1990, 2000; Humphrey 1992; Kornblith 1993). We are cognitively imprinted to discover, recognise and categorise certain patterns in the world – otherwise our generalisations and predictions would be miraculous (Dennett 1991; see also Johansson 1998). So, in contrast to the discussion mentioned above on what constitute laws, our focus concerning inductive inferences is on the kind of *knowledge* that is involved.

We show in section 6 that psychological research on sensory and perceptual generalisations involved in learning concern properties and categories, rather than propositions. From this perspective, inductive inferences can be seen as natural processes in cognitive systems, rather than in language, that occur when an agent categorises its sensory input and then makes generalisations or predictions using its understanding of these categories.

Our approach to induction is naturalistic in the sense that we look to science for relevant input instead of relying on intuitions and language.[1] Methodologically, we endorse a form of 'cooperative naturalism', according to which relevant scientific findings always should be taken into consideration since they provide our best explanations (Rysiew 2016).[2]

In fact, there are a number of interconnected scientific practices inquiring into induction and knowledge, on many different levels of explanation and from different perspectives. Three research areas come fairly close to the traditional epistemological outlook to induction, namely, cognitive neuroscience, cognitive ethology, and cognitive psychology. In section 4 we single out and use cognitive neuroscience as a foundation for our partitioning of knowledge types and in section 6 we turn to cognitive psychology for experimental evidential input concerning inductive reasoning.

3 Knowledge-how, knowledge-what and knowledge-that

3.1 The contemporary debate

Ryle (1949) provides some influential arguments for upholding the distinction between 'knowing-how' and 'knowing-that'. He argues that knowing-that is to *possess* knowledge whereas knowing-how is to be *intelligent*. Knowledge-that thus concerns relations between agents and true propositions, whereas knowledge-how instead concerns *abilities*, *dispositions* and *actions* of the agent.

However, not everyone agrees that there is a relevant distinction to be made (see, e.g., Stanley and Williamson 2001; Schaffer 2007; Stanley 2011). In particular, Stanley and Williamson (2001) and Stanley (2011) question the distinction and instead argue that knowledge-how is a form of knowledge-that. In the literature, this position is called *intellectualism*, in contrast to *anti-intellectualism* as exemplified by Ryle (1949), and for example Stanley (2011) claims that knowledge-how can be analysed as a state with propositional content.

In support of intellectualism, various examples like the following situation have been presented and discussed: "Suppose there is a certain complex ski manoeuver, which only the most physically gifted of athletes can perform. A ski instructor might know how to do that manoeuver, without being able to perform it herself." (Stanley 2011, p. 128).[3] The ski instructor is thought to know the relevant facts and propositions (knowledge-that) concerning the manoeuver, which then can be used to 'direct' her or someone else's actions. According to Stanley: "[… T]he acquisition of a skill is due to the learning of a fact [which] explains why certain acts constitute exercises of skill, rather than reflex. A particular action […] is a skilled action, rather than a reflex, because it is guided by knowledge […]" (Stanley 2011, p. 130).

[1] It should be pointed out that we consider philosophical questions important in their own right. Our point is that induction is not just a philosophical problem.

[2] Alternatives to our position can be found in for example 'replacement naturalistic' theories, where Quine (1969a) offers the most well-known account. Following a traditional understanding of Quine's position, epistemology should "simply fall […] into place as a chapter of psychology and hence of natural science." (Quine 1969a, p. 82). Yet another alternative position is found in 'substantial naturalism', according to which epistemological questions ought to be re-formulated in more exact scientific terminology (Rysiew 2016).

[3] Stanley attributes this example to p.c. with Jeff King.

The intellectualist position is, in our view, questionable since it fits only some aspects of highly technical skills and especially since it underestimates the importance and amount of non-conscious processes involved in intentional actions – even though there has been intellectualist attempts to better account for such aspects (see, e.g., Stanley and Krakauer 2013; Pavese 2015a, b). We agree that having propositional or theoretical knowledge (of true propositions), or receiving instructions ('knowledge of the way') that we should position and move our body in a particular manner might help us try to consciously improve our technique. Nevertheless, it is ultimately practical knowledge through repetitive training that eventually lets us *know how* to actually perform the action – it is only by going out on the slopes that we can learn how to ski. A myriad of non-declarative and non-conscious processes make up our motor-, perceptual- and cognitive abilities, which are required for us to know how to perform an action.[4]

3.2 Knowledge-what as knowledge of categories

Fantl proposes a more promising extension of the traditional dichotomy between knowledge-how and knowledge-that: "There's the kind of knowledge you have when it is truly said of you that you know a person—say, your best friend." (Fantl 2016). In our opinion, Fantl's knowledge of 'acquaintance' is a special case of a third type of knowledge. We want to single out the ability to *categorise*, in particular to know *the relation between categories and properties* as a special form of knowledge, which we call *knowledge-what*.[5]

Not all relations between categories and properties are, however, relevant for induction. To make our use of the term knowledge-what more precise, three types of information about categories need to be separated: Defining properties, characteristic properties and accidental facts (Keil and Batterman 1984).[6] Here we use these terms in the following way: *Defining* properties of a category refer to information that pertains to the meaning of the word for the category. *Characteristic* properties refer to general knowledge about the category, that is, properties that generally hold of the category (exceptions may be possible).[7] In the case when characteristic properties are formulated in sentences, the distinction between defining and characteristic corresponds to the distinction between definitional and law-like sentences that has been made within

[4] Another discussion of knowledge, which has received much less attention than knowledge-that and knowledge-how, is captured by the general formula *knowledge-wh*. This formula refers to the kind of knowledge involved when answering questions about who, when, where, why, whether, and what. If we consider knowledge-what, the examples that have been presented in the literature all concern singular facts rather than something general or categorical. Consequently, trying to identify the type of knowledge generated by induction by analysing answers to non-generic wh-questions does not seem to be a fruitful strategy. The intellectualist tradition claims that all forms of knowledge-wh, just as knowledge-how, reduces to declarative knowledge-that (Hintikka 1975; Lewis 1982; Boër and Lycan 1986; Higginbotham 1996; Stanley and Williamson 2001). We instead want to argue that knowledge-what is not directly connected to language but instead to properties and categories.

[5] It should be noted that our use of 'knowledge-what' is not intended to cover all everyday uses of the term such as in "I know what time it is".

[6] This is sometimes referred to as the dictionary-encyclopaedia distinction, but this is a misnomer since dictionaries frequently use characteristic features in their definitions.

[7] Among semanticists, it has been discussed how the borderline between defining and characteristic knowledge should be drawn, but this problem is not crucial for our arguments.

philosophy of science (Hempel 1965; Carroll 2016). *Accidental facts* contain information about particular instances of a category. We illustrate these three types of information with an example concerning the category 'spiders' web':

- *Defining*: Spiders' webs are made from a protein fibre extruded from the spider's body.
- *Characteristic*: Spiders' webs are used for catching insects that provide food for the spiders.
- *Accidental*: Spiders' webs are abundant in my cellar.

Our take on knowledge-what is that it concerns defining and characteristic knowledge, while knowledge-that concerns facts – accidental facts as well as facts of the type $2 + 2 = 4$. Our central thesis (to be discussed in section 6) is that, as a special case of knowledge-what, inductive inferences result in knowledge about characteristic properties. We thus heed the anti-intellectualist distinction between knowledge-how and knowledge-that while adding knowledge-what as a third type of knowledge, which is central for processes of induction.

3.3 Knowledge-what is separate from knowledge-that

Even though knowledge-what is primarily non-linguistic, it can be expressed in language. We next present two arguments for why knowledge-what, even if formulated linguistically, should be separated from knowledge-that. The first one builds on the observation that it seems perfectly natural that the following two sentences can be accepted simultaneously:

- (1) Spiders have eight legs.
- (2) My neighbour's spider has only seven legs.

However, if (1) is expressed – as is standard in writings on induction – in the form of a universal sentence:

- (1′) All spiders have eight legs.

Then (1′) contradicts (2). There are two reasonable ways out of the contradiction:

- (a) Deny that my neighbour's creature is a spider.
- (b) Deny that (1) expresses the same knowledge as (1′).

A reason against option (a) can be found in that what characterises spiders can be thought of as a 'pattern' of properties. Despite having only seven legs it is still a spider, since it has other 'essential' properties of spiders (definitional properties).[8] In favour of option (b), it is worth highlighting that (1) expresses definitional properties, while (2)

[8] We need not subscribe to full-blown essentialism. It is sufficient that certain properties of spiders are considered cognitively more important than others. See Gärdenfors (2000, sec. 4.2.2) for a defence of such a 'cognitive' form of essentialism.

expresses an accidental fact. Interpreting (1) as (1′) and putting it together with (2) conflates the two different types of knowledge. As an alternative way out of the contradiction one may propose the following formulation of (1):

- (1″) Spiders *characteristically* have eight legs.

Barring the problem of explaining the meaning of 'characteristically' in a non-circular way, we consider that this formulation supports our position that the knowledge expressed in (1) is of the definitional or characteristic form, that is, knowledge-what.

3.4 Generic sentences express knowledge-what

A second argument for maintaining the distinction between knowledge-what and knowledge-that shows up in natural language, albeit in an indirect way, as the distinction between the meaning of *generic* universals versus the meaning of factual universals. For example, generic universals such as "Blue whales eat plankton" and "A wrench is a tool for fastening nuts" are used to express some of the characteristic properties of 'whale' and 'wrench'. In contrast, factual universals, such as "Blue whales can be seen around the Cape of Good Hope" and "Wrenches are expensive in this shop" express facts about the world that are not part of the characteristic properties about the concepts. And sentence (1) above is indeed a generic.

It is interesting to note that the two types of universals behave in different ways linguistically, as pointed out by Lawler (1973):

- (3a) Blue whales eat plankton.
- (3b) A blue whale eats plankton.
- (4a) Blue whales can be seen around the Cape of Good Hope.
- (4b) *A blue whale can be seen around the Cape of Good Hope.[9]

(3a) describes a characteristic property of *blue whales*. It can be exchanged for the indefinite singular version in (3b). It expresses a relation between the concept blue whale and the property of feeding on plankton. In contrast, (4a) is a factual universal that says something factual about blue whales. A test for this is that it cannot be exchanged for the indefinite singular version in (4b) (Carlson 2009; Krifka 2012). Lawler notes that generic universals (which he calls non-descriptive generics) "[…] seem most natural in definitional sentences, or ones used somehow to identify the nature of the thing specified by the generic by means of properties peculiar to it; they are less acceptable when an accidental quality is predicated on them." (Lawler 1973, p. 112).

The upshot is that although a generic universal is a sentence, it expresses a different kind of knowledge than factual universals. Philosophers who have analysed generics have noted that there is no linguistic operator associated with these sentences and that negations of generics cannot be handled in the traditional logical way (Leslie 2008). The fact that sentences (3a) and (3b) express the same content in spite of their very different logical form is a further indication that generics form a special class of sentences. This conclusion is also supported by the fact that generics are acquired

[9] As is standard in linguistics, the asterix marks that the sentence is not acceptable.

earlier by children than explicit universal sentences (Gelman 2003), which indicates that the information contained in generics is of a more fundamental type (see also Hollander et al. 2002). Leslie (2008, p. 21) writes: “Thus the inclination to generalize, though aided by language, does not depend on language but is, rather, an early developing, presumably innate, cognitive disposition.”

Our position is that induction does not concern relations between sentences and hence it is not a logical problem involving relations between sentences. We submit that the focus should be on how relations between categories and properties are supported.

In this section we have argued that knowledge-how, knowledge-what and knowledge-that all fill important separate epistemic roles. We thus propose a tripartite division of knowledge. In the next section we present results from cognitive neuroscience that further support such a tripartition.

4 Memory and knowledge

Without memory there is no knowledge. In this section, we take a cognitive neuroscientific perspective and present a different kind of support for our thesis that knowledge-what is a separate form of knowledge by mapping our partitioning of the three types of knowledge onto different kinds of long-term memory. We build on Tulving’s (1985) categorisation of long-term memory into three kinds: procedural, semantic and episodic memories. Tulving’s position has been very influential and is still pertinent in recent analyses although it has been partially reinterpreted (see, e.g., Fletcher et al. 1999; Binder and Desai 2011; Yee et al. 2014; Kim 2016; see also Gazzaniga et al. 2002; Aizawa and Gillett 2009). In this paper, we follow the presentation in Yee et al. (2014).

4.1 Mapping forms of knowledge onto forms of memory

The non-declarative *procedural* memory, which is beyond our conscious reach, handles an agent’s skill in performing a task. This kind of memory can be described as generated by an automatic process, where an agent learns and remembers *how* to do something. Learning is achieved through repetition or practice, and procedural memory can easily be associated with operant conditioning since it is possible to describe in terms of stimulus and response. Procedural memory is something humans share with many other animals (Tulving 2002).

Semantic memory allows for agents to actively cognise about categories, concepts and objects. It is thus with the aid of semantic memory that agents think about categories and their relations (see, e.g., Herrnstein 1990; Martin et al. 1996; Martin and Chao 2001; Binder and Desai 2011; Yee et al. 2014). Semantic memory is general and does not depend on specific references to experiences. This kind of memory is needed for handling the environment as efficiently as possible. In particular, semantic memory is crucial for mapping categories to actions. Like procedural memory, some aspects of semantic memory are most likely hardwired through evolution, for example fear reactions to snakes.

> [C]ategorization is no saltation. It has turned up at every level of the animal kingdom where it has been competently sought. One reason for looking more

> carefully at lower levels of categorization is that the continuity of cognitive processes linking humans and other animals is clear and undeniable here. And, as the evidence to be summarized suggests, it is probably at the upper end of this span that animal and human cognitive capacities diverge. (Herrnstein 1990, p. 138)

Humans share semantic memory with mammals and birds (Tulving 2002). Numerous findings support conceptual and categorical abilities in animals such as for example common squirrel monkeys (Thomas and Kerr 1976), rhesus monkeys (Spaet and Harlow 1943; Sands et al. 1982; Schrier and Brady 1987), chimpanzees (Nissen 1953), and pigeons (Vetter and Hearst 1968; Zeiler 1969; Cerella 1979).[10]

Episodic memory governs experienced knowledge that can be used in narratives. This kind of memory generates self-aware *remembrance* concerning single events such as they are experienced from a first person perspective (Tulving 1985, 2002).

Episodic memory makes it possible for humans to 'time-travel' in their minds. It allows us to remember individual events or episodes and the order in which they have occurred. Tulving (2002) claims that this form of memory is only found in humans. This position has, however, recently been challenged by researchers in animal cognition (Clayton and Dickinson 1998; Gärdenfors and Osvath 2010; Osvath 2015) who argue that episodic memory, albeit to a limited extent, can be found in animals such as great apes and corvids.

The three systems are viewed as separate systems, although they most likely work in parallel, something Tulving acknowledges (see, e.g., Tulving 2002, p. 6; see also Yee et al. 2014). We now propose a straightforward mapping between the three kinds of knowledge and the three long-term memory systems: Procedural memory handles knowledge-how, semantic memory handles knowledge-what, and episodic memory handles knowledge-that.[11] Since the characterisation of the knowledge handled by the three memory systems clearly maps onto our description of the three kinds of knowledge, this mapping supports that the three types we distinguish indeed have different functions in human cognition.

From the perspective of this article, it is interesting to note that Tulving claims that the order in which memory types are presented here corresponds to the order in which they have emerged in the evolution of the animal world. In Tulving's words: "[…] Procedural memory entails semantic memory as a specialized subcategory, and […] semantic memory, in turn, entails episodic memory as a specialized subcategory." (Tulving 1985, pp. 2–3, italics removed). Both episodic and semantic memories therefore involve non-conscious aspects from procedural memory, which is prior. Since episodic memory is the memory-form most tightly connected with conscious experiences, thereby being connected to introspection and internalistic justification, it is no wonder that knowing-that is thought to be central for humans. However, for everyday problem solving and survival, the two other types are more essential. The fact that many animals have procedural and semantic memory while episodic memory is only well developed in humans indicates that, from an evolutionary point of view, knowing-how and knowing-what are more fundamental forms of knowledge than

[10] For a more exhaustive overview see Thompson (1995).

[11] In different articles, Tulving vacillates in his characterisation of memories of facts. Here we link factual knowledge to episodic, rather than to semantic, memory (see also Yee et al. 2014).

knowing-that. Our argument therefore supports an anti-intellectualist position. Rather than investigating how the concept ‘knowledge’ figures in language, our mapping between knowledge and long-term memory focuses on how humans, and other animals, actually use their knowledge as shown by different cognitive tasks.

4.2 Semantic memory and neuroscience

Neuroscientific results provide ample support for Tulving’s distinction between the three memory systems, holding the procedural, semantic and episodic memory systems separate. In particular, the left and right prefrontal cortices are considered to play a key role for separating semantic and episodic memory. There is however an on-going debate concerning the details of this separation (Fletcher et al. 1999, p. 176; Goel and Dolan 2000; Kim 2016). Semantic memory is connected to conceptual knowledge, and fMRI studies show that in addition to the prefrontal cortex, the anterior cingulate, the inferior parietal cortex, the thalamus, and the hippocampus are also to various degrees involved in categorisation (Goel and Dolan 2000; Grossman et al. 2002b). Furthermore, the prefrontal cortex and hippocampus are directly linked to inductive inferences (Goel and Dolan 2000; Grossman et al. 2002a; Hayes et al. 2010; Yee et al. 2014; Fisher et al. 2015). Such findings offer a non-linguistic backing of our tripartitioning of knowledge as well as our linking of conceptual knowledge and induction to semantic memory. Specific brain-regions are correlated to categorical, conceptual and inductive inferential functions, all being ascribed to semantic memory.

The credibility of our tripartite account of knowledge, given the evidence from neuroscience, offers a counterargument against a reduction of knowledge-what (or knowledge-how) into knowledge-that. Conceptual knowledge-what should not be seen as reducible to propositional knowledge-that, since the memory system underlying knowledge-what is different from that underlying knowledge-that.

The upshot is that the mapping between types of knowledge and long-term memory systems thus provides us with a naturalistic argument for separating knowledge-what from knowledge-that. Indeed, memory science endorses the view that long-term memory can be partitioned into three types, which supports our corresponding distinction between three types of knowledge. In brief, knowledge-what is a special form of knowledge, just as semantic memory is a special form of memory.

5 Using conceptual spaces to model knowledge-what

We claim that knowledge-what is a different type than knowledge-that. What except for language can be used to model knowledge-what?[12] In this section we propose that *conceptual spaces* (Gärdenfors 1990, 2000, 2014) is an appropriate tool for this task. This notion can be seen as a development of the ‘quality spaces’ in Quine (1960), the ‘attribute spaces’ in Carnap (1971) and the ‘logical spaces’ in Stalnaker (1981). In section 6 we then argue that conceptual spaces help us understand induction as a way of achieving knowledge-what.

[12] We cannot use the word ‘describe’ instead of ‘model’ since that would presuppose a linguistic approach to knowledge.

There exist several other models of categories and their relations apart from conceptual spaces, for example models based on prototypes in the tradition of Rosch (1975) or exemplar-based models (Nosofsky 1988). However, the focus on geometrical structure, in particular the use of convexity in representing categories, make conceptual spaces particularly well suited for handling inductive processes (Gärdenfors 1990, 2000).

5.1 Dimensions and domains

A conceptual space consists of a number of quality dimensions. Examples of quality dimensions are temperature, weight, brightness, pitch, and force, as well as the three ordinary spatial dimensions of height, width, and depth. Some quality dimensions are of an abstract non-sensory character.

The quality dimensions are grouped into *domains*. For example, the space domain consists of the dimensions width, depth and height, and the colour domain of the dimensions hue, saturation and brightness. The domains are described with the aid of different topological or metric structures. For example the ordinary space domain forms a 3-dimensional Euclidean space, the colour domain forms a double spindle (Gärdenfors 2000), and the domain of tonal harmony forms a torus (Shepard 1982).

The primary function of the domains is to represent various qualities of objects. Distances in the domains are inversely correlated to the *similarities* between properties. For example the distance between orange and red in the colour domain is smaller than the distance between red and green. The domains of a conceptual space are related in various ways, since the properties of those objects modelled in the space co-vary. For example, in the fruit domain, the ripeness and colour dimensions co-vary and, of course, size and weight covariate strongly. Such covariations are central to inductive inferences.

The conceptual space framework presented here could be the answer to what for example Yee et al. (2014) are looking for in a domain-specific framework for semantic memory:

> Many of the studies described in this chapter explored the organization of semantic memory by comparing the neural responses to traditionally defined categories (e.g., animals vs. tools). However, a more fruitful method of understanding conceptual representations may be to compare individual concepts to one another, and extract dimensions that describe the emergent similarity space. (Yee et al. 2014, p. 363)

5.2 Conceptual spaces as a tool for expressing properties, categories, and their relations

In first-order logic and other logical formalisms, *properties* are described with the aid of predicates. However, predicates are treated as atoms and not further analysed. In contrast, if conceptual spaces are used to define properties, more structure can be represented. The central role of similarity and the geometry of the spaces make it possible to represent features of concepts and their relations that are more or less impossible to express within a logical approach (that is, as part of knowledge-that).

The following criterion was proposed in Gärdenfors (1990, 2000), where the geometrical characteristics of the quality dimensions are used to introduce a spatial structure to properties:

- *Criterion P*: A *natural property* is a convex region in some domain.

That a region is convex means that, if some objects located at x and y in relation to some domain are both examples of a property, then any object that is located *between* x and y with respect to the same domain will also be an example of the property. As an application of criterion *P*, Jäger (2010) has provided strong support for the convexity of colour terms in 109 languages. Properties as defined in this criterion are natural in the sense that they emerge as results of learning in children, adults and many animal species. We will discuss the relations to learning further in section 6.

The notion of a natural property can also be extended to some discrete dimensions. For example, in a graph structure with nodes and arcs, we have a notion of betweenness, and thus we can identify the convex sub-sets of the graph (compare Johnson's (1921, pp. 181–3) notion of 'adjectival betweenness'). This means that in a biological classification, which can be represented by a tree structure, a property is 'natural' if it applies to all and only those parts of the classificatory tree that lie below one particular node in the tree. For example, the properties 'marsupial' and 'vertebrate' will be natural properties in the phylogenetic classification, while 'featherless' and 'biped' will not.

Properties, as defined by criterion *P*, should be distinguished from *categories*. Gärdenfors (2000, 2014) defines this distinction by saying that a property is based on a single domain, while a category is based on one *or more* domains. This distinction has been obliterated in the philosophical literature since both properties and categories are represented by predicates in first-order logic. A rule of thumb is that adjectives in a language typically express properties, while nouns express categories. This point is developed in Gärdenfors (2014).

When representing a category, one of the first problems one encounters is to decide which the relevant domains are. A typical example of a category that is represented in several domains is 'apple' (compare Smith et al. 1988). When we encounter apples as children, the first domains we learn about are those of colour, shape, texture, and taste (see, e.g., Son et al. 2008; Gärdenfors 2017). Later, we learn about apples as fruits (biology), and as things with nutritional value, etc.

Categories are not just bundles of properties. They also involve relations and *covariations* between regions from different domains that are associated with the category. The 'apple' category has a strong positive covariation between sweetness in the taste domain and sugar content in the nutrition domain, and a weaker positive covariation between redness and sweetness. Such considerations motivate the following definition for category[13]:

- *Criterion C*: A *natural category* is represented as a set of convex regions in a number of domains, together with information about how the regions in different domains are related.

[13] For a more precise definition, see Gärdenfors (2000, ch. 4).

Within the philosophical tradition, a forerunner is Johnson's (1921, ch. XI) distinction between 'complex determinables' (corresponding to our domains) and 'determinates' (corresponding to points or regions of domains).

The theory of conceptual spaces has clear connections to prototype theory, according to which members in each category in a domain are more or less typical. The member that is most typical can be dubbed a prototype, although it can be pointed out that properties often do not have clear-cut lines but instead graded boundaries (Rosch 1975; Smith et al. 1988; Decock et al. 2013). This is easily translated into the terminology of conceptual spaces where a prototype can be described as lying at the centre of the region(s) representing a property or category.

One aspect that deserves to be highlighted is that conceptual spaces offer the ability to add domains to the representation of a concept (Gärdenfors 2000, 2104). To use our previous example, when we learn the meaning of 'apple' as children, the shape, colour and taste domains are the central ones. Later we learn that apples also have nutritional values, which can be represented by adding a new domain to the 'apple' category. Adding new domains is a form of *learning* about categories.

The connections to prototype theory and the possibility to learn about a category by adding new domains entails that our model of properties and categories is in conflict with the classical approach where concepts are defined in terms of necessary and sufficient conditions. The classical approach presumes a language-based description of concepts, something that is not presupposed when concepts are represented in terms of conceptual spaces.

Our interpretation of conceptual spaces is instrumentalistic. Nevertheless, our evolutionarily moulded cognitive faculties provide some natural quality dimensions for humans. Our quality dimensions are what they are because they have been selected to fit the surrounding world (Gärdenfors 2000, p. 82). In Quine's words: "To trust induction as a way of access to the truths of nature [...] is to suppose, more nearly, that our quality space matches that of the cosmos." (Quine 1969b, p. 125). His notion of 'quality space' is close to that of a conceptual space.

5.3 Knowledge-what as relations between categories and properties

We might not have unbiased contact with the world, but it is still the real world that provides the sensory input we get, and "[i]t is precisely because the world has the causal structure required for the existence of natural kinds that inductive knowledge is even possible." (Kornblith 1993, p. 35). There are only certain clusters of properties that are organised in a stable enough way as to stick together in natural categories enabling us to make inductive inferences.

As a way of capturing clusters of properties, criterion *C* introduces relations between domains as a factor of an object category. Our proposal is that knowledge-what consists of such relations. For example knowing what aspartame is, involves knowledge about the relation between the chemical domains that characterise aspartame and the sweetness region of the taste domain.[14]

There are, however, different kinds of relations. The strongest one is when all examples of a category fall within one region of a domain, as in for example "all

[14] See Gärdenfors (2000) for a discussion of this domain.

ravens are black". Another form of relation is *covariation*, for example, "metals expand when heated" which describes a covariation between the temperature and size domain, or the covariation between the colour and sweetness of fruits.

Even though there are many possible relations between categories and properties, we most often are able to discern which relations are relevant. We have a built-in understanding of the world's structure, where some properties and patterns are intuitively grasped (Kornblith 1993, pp. 100–1; Johansson 1998). As Kornblith points out "[…] we are accomplished detectors of multiple, clustered patterns of covariation." (Kornblith 1993, p. 104). It is primarily in contrived situations that our inductive inferences tend to go wrong; in natural settings we are quite apt at recognising essential 'deep similarities':

> It is thus safe to say that we have a sensitivity to the features of objects which reside in homeostatic clusters. Indeed, the way in which we detect covariation is precisely tailored to the structure of natural kinds. [… W]e conceptualize kinds in such a way in order to separate the properties of the members of a kind which are projectable from those which are not. We are aided in this task by our ability to detect clustered covariation. (Kornblith 1993, pp. 105–6)

An argument for focusing on covariations between domains comes from work by Billman (1983) and Billman and Knutson (1996) that indicates that humans are quite good at detecting covariations that cluster several domains (Hayes et al. 2010). A plausible explanation of this phenomenon is that our perceptions of natural objects show covariations along multiple domains, and, as a result of natural selection, we have developed a competence to detect such clustered relations. In line with this, the *basic level* categories of prototype theory (Rosch 1975) can be characterised by distinctive clusters of covariating properties (Holland et al. 1986, pp. 183–4).

6 Induction as generating knowledge-what

As we mentioned earlier, the propositional approach to induction led to unintuitive conclusions visible in numerous paradoxes. Quine's (1969b) negative conclusions concerning the possibilities of defining 'natural kind' or the corresponding notion 'similarity' can be interpreted as indicating that we have to go beyond language to find a solution. What is needed is a way of tapping our sources of knowledge so that we become able to distinguish the properties that may be used in inductive inferences from those that may not.

For Goodman (1983), the question of what makes certain generalisations law-like becomes the problem of which predicates are 'projectable', that is, which predicates can be used in inductive inferences.[15] The solution we propose here is that only natural properties and categories, as defined in criteria *P* and *C*, are 'projectable', that is, allowed in inductive inferences (Gärdenfors 1990, 2000). Consequently, the feature of a conceptual space that is the most essential for a theory of induction is its topological

[15] Goodman spells the term 'projectible' in his well-known discussion of induction and entrenched predicates (Goodman 1983).

and metric properties, while the logical structure of the language that 'lives on' in the conceptual space is secondary.

As an example, let us take a brief look at the categories that occur in Hempel's (1965) paradox of confirmation. The paradox describes how all observations of black ravens confirm the generalisation that all ravens are black. However, all observed non-black non-ravens logically confirm the same generalisation, which might be considered counterintuitive or paradoxical. Observing a white shoe, for example, would confirm that all ravens are black. It seems odd that any such observation should support an inductive inference that all ravens are black. According to the theory of Gärdenfors (2000, 2014), object categories are represented in product spaces, where each subspace represents a property of the category. If the properties of the category all correspond to convex regions, then the product of the regions will be convex too.[16] In contrast, the category 'non-raven' would be difficult to count as a natural category. The class of all objects that are non-ravens belong to many unrelated domains. The associated regions, let alone their product, cannot be specified as a convex region of some domain. Consequently, 'non-raven' does not qualify as a natural category. A similar analysis can be provided for Goodman's (1983) example of 'grue' (Gärdenfors 1990). The properties used in these problems do not correspond to convex regions in domains and are hence not projectable. A more detailed discussion of this topic is presented in Gärdenfors (2000).

Conceptual spaces help us understand induction as a way of achieving knowledge-what. Given the characterisation of projectable predicates as natural properties and categories, our analysis of what is achieved in induction is knowledge about relations between such properties and categories. Induction thus lets us achieve new knowledge-what about categories. And to know what properties a specific category is related to is to have knowledge about characteristic properties of that category. Most scientific empirical discoveries are of this type. For example, when it was discovered that penicillin is an antibiotic, such a relation was established (Aldridge et al. 1999; Lax 2004). Or when it was discovered that a certain alloy of niobium and titanium was a superconductor (Berlincourt and Hake 1963) new knowledge about characteristic properties of the type knowledge-what was acquired.

Induction consists in generalising from a limited number of observations. In logical approaches to induction, 'generalisation' means forming some form of universal sentence. When conceptual spaces are used as a basis, however, the situation is different. The similarity structure of the domains allow generalisation in the form of extending the given observations to *similar* instances, in particular by applying the convexity criteria *P* and *C*. This form of generalisation therefore comes closer to what is called 'stimulus generalisation' in psychology. Unfortunately, this form of generalisation has not been discussed in the philosophical literature on induction.

In contrast to the sentential approach to induction in philosophy, there exist in psychology an active research programme dealing with 'category-based induction' (Osherson et al. 1990; Hayes et al. 2010; Fisher et al. 2015). Within this programme, the stimuli almost exclusively consist of generic sentences that, according to our classification, express knowledge-what. The inferences that are studied are typically

[16] The model presented in Gärdenfors (2000, 2014) is slightly more complicated, involving also correlations between properties.

of two kinds: general, where the conclusion concerns a class that is superordinate to those of the premises, and specific, where the class of the conclusion is on the same categorical level as the premises. An example of a general argument is the following:

- Grizzly bears love onions
- Polar bears love onions

Hence: All bears love onions.

And an example of a specific argument is:

- Robins use serotonin as a neurotransmitter
- Bluejays use serotonin as a neurotransmitter

Hence: Geese use serotonin as a neurotransmitter.

Experimental subjects are asked to judge the validity of different inductive relations.[17] A central question that is investigated is how the perceived similarities between the categories affect the judgments. Thus category-based induction is closely related to stimulus generalisation. In accordance with our analysis, the focus of this research programme is knowledge-what. As far as we are aware the distinction between knowledge-what and knowledge-that has not been discussed within this psychological tradition.

Further support for the thesis that inductive generalisations build on relations between categories comes from studies of how children reason (Sutherland and Cimpian 2017). It has been argued that the drive to learn about categories is an innate feature of human cognition (Csibra and Gergely 2009). Information about categories is privileged in memory since children are better able to recall new information about categories than to recall information about non-category sets (Cimpian and Erickson 2012). Furthermore, children find it easier to reason with categories (dogs) than with set-expressions (all dogs) (Hollander et al. 2002). Findings of this type indicate that knowledge-what is primary to knowledge-that, just as semantic memory is primary to episodic.

It should be noted that the change in how induction is perceived – from relations between sentences to relations between properties and categories – does not lead to any radical changes in the *methodology* used to establish inductive knowledge. Well-known requirements of repeated experiments, precision, variation and generalisability in experiments are still valid (Hempel 1965; Seltman 2015). Furthermore, these requirements turn out to be even more natural from the perspective of establishing knowledge-what. As noted above, generalisability achieves a different meaning and different methods for determining relations that take distances in domains into consideration should therefore be put in focus.

7 Concluding remarks

We have argued that the traditional problems for the logical positivists' analyses of induction have arisen because they confined themselves to narrowly to sentential

[17] When we write about validity, we intend not just logical validity, but use the term in a broader sense including other forms of inference.

representations of information and to logical tools in their analyses. Instead we have shown the fruitfulness of using conceptual spaces as a way to represent knowledge-what and to investigate inductive inferences.

We have defended two theses. Firstly, there is not only knowledge-how and knowledge-that, but also knowledge-what. Secondly, induction concerns knowledge-what, that is, knowledge concerning the relation between categories and properties. We have presented support for these theses by connecting our tripartition of knowledge to the procedural, semantic and episodic long-term memory systems. We have specifically stressed the correlations in brain activity found between semantic memory, conceptual knowledge and induction. Knowledge-what should thus be included as a fundamental component of an account of human knowledge.

It is time to give up the focus on propositional knowledge in analytical philosophy. In our opinion, the many riddles of induction are a consequence of this focus and they will not appear if knowledge-what is accepted as a type of knowledge and induction is recognised as involving knowledge-what. By introducing the tripartition of knowledge, we hope to reboot epistemology in a naturalistic direction. Since philosophical (but not psychological) research on inductive processes during the last century has focused on symbolic representations of knowledge-that, we propound that the representations of categories and properties – as a way of modelling knowledge-what – should be given much more attention in the future.

Acknowledgements We want to thank Ingvar Johansson, Jan Faye, and our anonymous reviewers for valuable comments on earlier version of the text. Peter Gärdenfors gratefully acknowledges support from the Swedish Research Council to the Linneaus environment Cognition, Communication and Learning. He also thanks Wissenschaftskolleg zu Berlin for support during the writing of this paper.

References

Aizawa, K., & Gillett, C. (2009). The (multiple) realization of psychological and other properties in the sciences. *Mind & Language, 24*(2), 181–208. https://doi.org/10.1111/j.1468-0017.2008.01359.x.

Aldridge, S., Parascandola, J., & Sturchio, J. L. (1999). *Discovery and development of penicillin*. American Chemical Society International Historic Chemical Landmarks. http://www.acs.org/content/acs/en/education/whatischemistry/landmarks/flemingpenicillin.html/ .

Armstrong, D. M. (1978). *A theory of universals*. Cambridge: Cambridge University Press.

Armstrong, D. M. (1983). *What is a law of nature*. Cambridge: Cambridge University Press.

Berlincourt, T. G., & Hake, R. R. (1963). Superconductivity at high magnetic fields. *Physical Review, 131*(1), 140–157. https://doi.org/10.1103/PhysRev.131.140.

Billman, D. (1983). *Procedures for learning syntactic structure: A model and west with artificial grammars*. Doctoral dissertation: University of Michigan.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 458–475.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001.

Boër, S., & Lycan, W. (1986). *Knowing who*. Cambridge: Cambridge University Press.

Carlson, G. N. (2009). Generics and concepts. In F. J. Pelletier (Ed.), *Kinds, things and stuff: Mass terms and generics* (pp. 16–36). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195382891.003.0002.

Carnap, R. (1950). *Logical foundations of probability.* Chicago, IL: Chicago University Press.

Carnap, R. (1971). A basic system of inductive logic, part 1. In R. Carnap & R. C. Jeffrey (Eds.), *Studies in inductive logics and probability* (Vol. 1, pp. 35–165). Berkeley: University of California Press.

Carroll, J. W. (2016). Laws of nature. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2016 Edition)*. URL = http://plato.stanford.edu/archives/fall2016/entries/laws-of-nature/

Cerella, J. (1979). Visual classes and natural categories in the pigeon. *Journal of Experimental Psychology: Human Perception and Performance, 5*(1), 68–77.

Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: New evidence that categories are privileged in children's thinking. *Cognitive Psychology, 64*(3), 161–185. https://doi.org/10.1016/j.cogpsych.2011.11.002.

Clayton, N. S., & Dickinson, A. D. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature, 395*(6699), 272–274. https://doi.org/10.1038/26216.

Creath, R. (2014). Logical empiricism. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2014 Edition)*. URL = http://plato.stanford.edu/archives/spr2014/entries/logical-empiricism/

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005.

Decock, L., Dietz, R., & Douven, I. (2013). Modelling comparative concepts in conceptual spaces. In Y. Motomura, A. Butler, & D. Bekki (Eds.), *New Frontiers in artificial intelligence* (pp. 69–86). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39931-2_6.

Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy, 88*(1), 27–51. https://doi.org/10.2307/2027085.

Dretske, F. (1977). Laws of nature. *Philosophy of Science, 44*(2), 248–268. https://doi.org/10.1086/288741.

Fantl, J. (2016). Knowledge how. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition)*. URL = https://plato.stanford.edu/archives/spr2016/entries/knowledge-how/

Fisher, A. V., Godwin, K. E., Matlen, B. J., & Unger, L. (2015). Development of category-based induction and semantic knowledge. *Child Development, 86*(1), 48–62. https://doi.org/10.1111/cdev.12277.

Fletcher, P. C., Büchel, C., Josephs, O., Friston, K., & Dolan, R. J. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex, 9*(2), 168–178. https://doi.org/10.1093/cercor/9.2.168.

Gärdenfors, P. (1990). Induction, conceptual spaces and AI. *Philosophy of Science, 57*(1), 78–95. https://doi.org/10.1086/289532.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge: Bradford Books, MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge: MIT Press.

Gärdenfors, P. (2017). Semantic knowledge, domains of meaning and conceptual spaces. In P. Meusburger, B. Werlen, & L. Suarsana (Eds.), *Knowledge and action. Knowledge and space* (Vol. 9, pp. 203–219). Cham: Springer. https://doi.org/10.1007/978-3-319-44588-5_12.

Gärdenfors, P., & Osvath, M. (2010). Prospection as a cognitive precursor to symbolic communication. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *Evolution of language: Biolinguistic approaches* (pp. 103–114). Cambridge: Cambridge University Press.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive neuroscience: The biology of the mind*. New York: W. W. Norton & Company.

Gelman, S. A. (2003). *The essential child*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195154061.001.0001.

Goel, V., & Dolan, R. J. (2000). Anatomical segregation of component processes in an inductive interference task. *Journal of Cognitive Neuroscience, 12*(1), 110–119. https://doi.org/10.1162/08989290051137639.

Goodman, N. (1983, [1955]). *Facts, fiction and forecast* (4th ed.). Cambridge: Harvard University Press.

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., & Gee, J. (2002a). The neural basis for category-specific knowledge: An fMRI study. *NeuroImage, 15*(4), 936–948. https://doi.org/10.1006/nimg.2001.1028.

Grossman, M., Smith, E. E., Koenig, P., Glosser, G., DeVita, C., Moore, P., & McMillan, C. (2002b). The neural basis for categorization in semantic memory. *NeuroImage, 17*(3), 1549–1561. https://doi.org/10.1006/nimg.2002.1273.

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(2), 278–292. https://doi.org/10.1002/wcs.44.

Hempel, C. G. (1965). *Aspects of scientific explanation, and other essays in the philosophy of science*. New York: Free Press.

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition, 37*(1), 133–166. https://doi.org/10.1016/0010-0277(90)90021-B.

Higginbotham, J. (1996). The semantics of questions. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 361–383). Oxford: Blackwell Publishers Ltd.

Hintikka, J. (1975). Different constructions in terms of the basic epistemological verbs: a survey of some problems and proposals. In *The intensions of intentionality and other new models for modalities* (pp. 1–25). Dordrecht: D. Reidel.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge: MIT Press.

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology, 38*(6), 883–894. https://doi.org/10.1037/0012-1649.38.6.883.

Hume, D. (1988 [f.p. 1748]). *An enquiry concerning human understanding*. Illinois: Open Court.

Humphrey, N. (1992). *A history of the mind: Evolution and the birth of consciousness*. New York: Copernicus, Springer-Verlag. https://doi.org/10.1007/978-1-4419-8544-6.

Jäger, G. (2010). Natural color categories are convex sets. *Amsterdam Colloquium 2009, LNAI 6042*, 11–20.

Johansson, I. (1998). Pattern as an ontological category. In N. Guarino (Ed.), *Formal ontology in information systems* (pp. 86–94). Amsterdam: IOS Press.

Johnson, W. E. (1921). *Logic, part I*. Cambridge: Cambridge University Press.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 221–236. https://doi.org/10.1016/S0022-5371(84)90148-8.

Kim, H. (2016). Default network activation during episodic and semantic memory retrieval: A selective meta-analytic comparison. *Neuropsychologia, 80*, 35–46. https://doi.org/10.1016/j.neuropsychologia.2015.11.006.

Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge: MIT Press.

Krifka, M. (2012). Definitional generics. In A. Mari, C. Beyssade, & F. Prete (Eds.), *Genericity* (pp. 372–389). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199691807.003.0015.

Ladyman, J. (2002). *Understanding philosophy of science*. London: Routledge. https://doi.org/10.4324/9780203463680.

Lawler, J. M. (1973). Studies in English generics. (Doctoral dissertation, University of Michigan). University of Michigan Papers in Linguistics.

Lax, E. (2004). *The mold in Dr. Florey's coat: The story of the penicillin miracle*. New York: Holt Paperbacks.

Leslie, S. J. (2008). Generics: Cognition and acquisition. *Philosophical Review, 117*(1), 1–47. https://doi.org/10.1215/00318108-2007-023.

Lewis, D. (1982). Whether report. In T. Pauli (Ed.), *Philosophical essays dedicated to Lennart Åqvist on his fiftieth birthday* (pp. 194–206). Uppsala: Filosofiska studier.

Lorenz, K. (1977 [f.p. 1973]). *Behind the mirror: A search for a natural history of human knowledge*. London: Methuen.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology, 11*(2), 194–201. https://doi.org/10.1016/S0959-4388(00)00196-3.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*(6566), 649–652. https://doi.org/10.1038/379649a0.

Mill, J. S. (1843). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation* (Vol. *1*). London: John W. Parker.

Nissen, H. W. (1953). Sensory patterning versus central organization. *The Journal of Psychology, 36*(2), 271–287. https://doi.org/10.1080/00223980.1953.9712893.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 700–708.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185–200. https://doi.org/10.1037/0033-295X.97.2.185.

Osvath, M. (2015). Putting flexible animal prospection into context: Escaping the theoretical box. *WIREs Cognitive Science 2015*. https://doi.org/10.1002/wcs.1372, 1, 5, 18.

Pavese, C. (2015a). Knowing a rule. *Philosophical Issues: A Supplement to Nous, 25*(1), 165–188. https://doi.org/10.1111/phis.12045.

Pavese, C. (2015b). Practical Senses. *Philosopher's Imprint, 15*(29), 1–25.

Peirce, C. S. (1955 [f.p. 1883]). The general theory of probable inference. In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 190–217). New York: Dover Publications, Inc.

Quine, W. V. O. (1960). *Word and object*. Cambridge: MIT Press.

Quine, W. V. O. (1969a). Epistemology naturalized. In W. V. O. Quine, *Ontological relativity and other essays* (pp. 69–90). The John Dewey Essays in Philosophy, No. 1. New York: Columbia University Press.

Quine, W. V. O. (1969b). Natural kinds. In W. V. O. Quine, *Ontological relativity and other essays* (pp. 114–138). The John Dewey Essays in Philosophy, No. 1. New York: Columbia University Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192.

Rosenberg, A. (2000). *Philosophy of science: A contemporary introduction* (2nd ed.). New York: Routledge.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Rysiew, P. (2016). Naturalism in epistemology. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2016 Edition)*. URL = http://plato.stanford.edu/archives/sum2016/entries/epistemology-naturalized/

Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *Journal of Experimental Psychology: General, 111*(4), 369–389. https://doi.org/10.1037/0096-3445.111.4.369.

Schaffer, J. (2007). Knowing the answer. *Philosophy and Phenomenological Research, 75*(2), 383–403. https://doi.org/10.1111/j.1933-1592.2007.00081.x.

Schrier, A. M., & Brady, P. M. (1987). Categorization of natural stimuli by monkeys (*Macaca Mulatta*): Effects of stimulus set size and modification of exemplars. *Journal of Experimental Psychology: Animal Behavior Processes, 13*(2), 136–142.

Seltman, H. J. (2015). *Experimental design and analysis*. Pittsburgh: Carnegie Mellon University.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review, 89*(4), 305–333. https://doi.org/10.1037/0033-295X.89.4.305.

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12*(4), 485–527. https://doi.org/10.1207/s15516709cog1204_1.

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition, 108*(3), 626–638. https://doi.org/10.1016/j.cognition.2008.05.002.

Spaet, T., & Harlow, H. F. (1943). Solution by rhesus monkeys of multiple sign problems utilizing the oddity technique. *Journal of Comparative Psychology, 35*(2), 119–132. https://doi.org/10.1037/h0059354.

Stalnaker, R. (1981). Antiessentialism. *Midwest Studies of Philosophy, 4*, 343–355.

Stanley, J. (2011). *Know how*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199695362.001.0001.

Stanley, J., & Krakauer, J. W. (2013). Motor skill depends on knowledge of facts. *Frontiers in Human Neuroscience, 7*(503). https://doi.org/10.3389/fnhum.2013.00503.

Stanley, J., & Williamson, T. (2001). Knowing how. *The Journal of Philosophy, 98*(8), 411–444. https://doi.org/10.2307/2678403.

Sutherland, S. L., & Cimpian, A. (2017). Inductive generalization relies on category representations. *Psychonomic Bulletin and Review*, *24*(2), 632–636. https://doi.org/10.3758/s13423-015-0951-z .

Thomas, R. K., & Kerr, R. S. (1976). Conceptual conditional discrimination in *Saimiri Sciureus*. *Animal Learning & Behavior, 4*(3), 333–336. https://doi.org/10.3758/BF03214060.

Thompson, R. K. (1995). Natural and relational concepts in animals. In H. L. Roitblad & J. A. Meyer (Eds.), *Comparative approaches to cognitive science* (pp. 175–224). Cambridge: MIT Press.

Tooley, M. (1977). The nature of laws. *Canadian Journal of Philosophy, 7*(4), 667–698. https://doi.org/10.1080/00455091.1977.10716190.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1–12. https://doi.org/10.1037/h0080017.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114.

Vetter, G., & Hearst, E. (1968). Generalization and discrimination of shape orientation in pigeons. *Journal of the Experimental Analysis of Behavior, 11*(6), 753–765. https://doi.org/10.1901/jeab.1968.11-753.

Vickers, J. (2014). The problem of induction. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2014 Edition)*. URL = http://plato.stanford.edu/archives/fall2014/entries/induction-problem/

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2014). Semantic memory. In K. Ochsner & S. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience: Volume 1, core topics* (pp. 353–374). Oxford: Oxford University Press.

Zeiler, M. (1969). Repeated measurements of reinforcement schedule effects on gradients of stimulus control. *Journal of the Experimental Analysis of Behavior, 12*(3), 451–461. https://doi.org/10.1901/jeab.1969.12-451.

Paper III

A. Stephens
Three levels of naturalistic knowledge
Conceptual spaces: Elaborations and applications, 2019, Vol. 405, pp. 57-73

Chapter 4
Three Levels of Naturalistic Knowledge

Andreas Stephens

Abstract A recent naturalistic epistemological account suggests that there are three nested basic forms of knowledge: procedural knowledge-how, conceptual knowledge-what, and propositional knowledge-that. These three knowledge-forms are grounded in cognitive neuroscience and are mapped to procedural, semantic, and episodic long-term memory respectively. This article investigates and integrates the neuroscientifically grounded account with knowledge-accounts from cognitive ethology and cognitive psychology. It is found that procedural and semantic memory, on a neuroscientific level of analysis, matches an ethological reliabilist account. This formation also matches System 1 from dual process theory on a psychological level, whereas the addition of episodic memory, on the neuroscientific level of analysis, can account for System 2 on the psychological level. It is furthermore argued that semantic memory (conceptual knowledge-what) and the cognitive ability of categorization are linked to each other, and that they can be fruitfully modeled within a conceptual spaces framework.

Keywords Naturalistic epistemology · Cognitive philosophy · Conceptual knowledge · Knowledge-what · Categorization · Conceptual spaces

4.1 Introduction

Investigations regarding knowledge have been going on for millennia while the concept still lacks a sharp and widely accepted definition (see, e.g., Markie 2013; Samet and Zaitchik 2014). However, many philosophers nowadays heed naturalism and consider it the job of science to provide our best explanations. Furthermore, as cognitive sciences have progressed, much relevant information regarding our cognitive faculties and knowledge is indeed available. We understand the world through multiple models, but since different sciences explore cognition

A. Stephens (✉)
Lund University, Lund, Sweden

M. Kaipainen et al. (eds.), *Conceptual Spaces: Elaborations and Applications*,
Synthese Library 405, https://doi.org/10.1007/978-3-030-12800-5_4

and knowledge on different levels of analysis, it is not clear if, or how, the different accounts of knowledge they provide can, or should, be united (see, e.g., Dupré 1993; Mitchell 2003; Horst 2016).

In an attempt to offer some clarity and coherence, Gärdenfors and Stephens (2018) have argued that there are three nested basic forms of knowledge: procedural knowledge-how, conceptual knowledge-what, and propositional knowledge-that. The tri-partite knowledge-account is grounded in cognitive neuroscience where the three forms of knowledge are mapped to procedural, semantic, and episodic memory respectively. While there is an extensive and on-going epistemological discussion concerning the traditional forms knowledge-how and knowledge-that (see, e.g., Ryle 1949; Stanley 2011; Fantl 2016), a lot remains to be explored regarding the form knowledge-what, which Gärdenfors and Stephens argue is generated by inductive reasoning.

Moreover, in encouragement of a multi-disciplinary and multi-level development of our understanding of knowledge, cognition and behavior (see, e.g., Frank and Badre 2015), it can be pointed out that:

> [T]he neurosciences are reshaping the landscape of the behavioral sciences, and the behavioral sciences are of increasing importance to the neurosciences, especially for the rapidly expanding investigations into the highest level functions of the brain. (Berntson and Cacioppo 2009, p. xi)

This article attempts to broaden the proposed knowledge-account and our understanding of knowledge-what by investigating two issues. First, the prospect of integrating the knowledge-account with models from two other scientific perspectives (cognitive ethology and cognitive psychology) on higher levels of analysis will be explored. If successful, such integration would increase the knowledge-account's plausibility. By encompassing three levels of analysis, it would present a naturalistic framework arguably fairly close to a traditional epistemological outlook. Second, the link between the knowledge-form knowledge-what and *categorization* will be considered. I will loosely follow a prototype theoretical interpretation and view categorizations as natural cognitive phenomena where organisms try to acquire as much information as possible of the surrounding structured world, while minimizing their energy-expenditure (see, e.g., Rosch 1975a, b). According to such an interpretation, objects in a category are compared in relation to how representable they are, and the most representable object is seen as a prototype. Other objects can then be compared in relation to how similar they are to the prototype (see, e.g., Gärdenfors 2000, p. 84). This inquiry diverges from Gärdenfors and Stephens' discussion, which centers on the specific role of inductive inferences, and can thus be seen as a complementary development of their account.

After this short introduction Sect. 4.2 will give an outline of Gärdenfors and Stephens' knowledge-account grounded in cognitive neuroscience. Section 4.3 then investigates knowledge from the perspective of cognitive ethology, and the possibility of integrating the cognitive ethological account with the neuroscientific account. Section 4.4 continues by inquiring into how a cognitive psychological account can be integrated with both former accounts, and in Sect. 4.5 it is lastly argued that conceptual knowledge-what and categorizations can be fruitfully modeled within a conceptual spaces framework (Gärdenfors 1990, 2000, 2014).

4.2 Cognitive Neuroscience: Knowledge and Memory

Even though there are various different models and theories pertinent to understand knowledge from a neuroscientific proximate perspective, Gärdenfors and Stephens (2018) single out and use Tulving's (1985; see also 1972) seminal account of memory and consciousness. This is, arguably, a reasonable basis since Tulving's account has been extremely influential and is often used as a starting-point in neuroscientific research even by those who ultimately deviate from it. Knowledge, from a neuroscientific perspective, is thought to have its foundation in long-term memory (LTM), and Tulving divides LTM into three nested parts, illustrated in Fig. 4.1: procedural memory, semantic memory, and episodic memory, where '[. . .] procedural memory entails semantic memory as a specialized subcategory, and in which semantic memory, in turn, entails episodic memory as a specialized subcategory.' (Tulving 1985, pp. 2–3, italics removed; see also Fletcher et al. 1999; Goel and Dolan 2000; Kan et al. 2009; Barrett 2015; Kim 2016). Tulving argues that:

> Procedural memory [. . .] is concerned with how things are done – with the acquisition, retention, and utilization of perceptual, cognitive, and motor skills. Semantic memory – also called generic [. . .] or categorical memory [. . .] – has to do with the symbolically representable knowledge that organisms possess about the world. Episodic memory mediates the *remembering* of personally experienced events [. . .]. (Tulving 1985, p. 2)

With this partitioning as an underpinning, and trailing the neuroscientific canon, procedural knowledge-how (the knowledge of *how* to ride a bike – an ability) readily

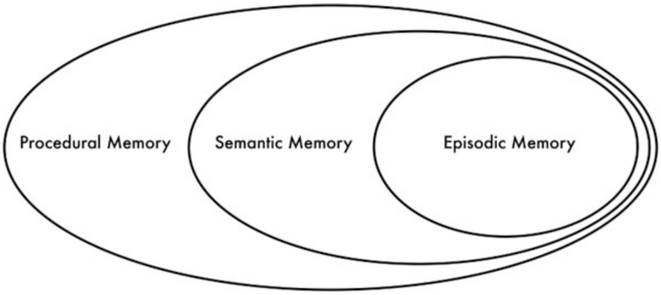


Fig. 4.1 Tulving's nested account of the LTM. Procedural memory entails semantic memory as a specialized subcategory, and semantic memory entails episodic memory as a specialized subcategory

maps to non-declarative procedural memory. This form of memory governs actions while it to a large extent is automatic and non-conscious. Through repetition we can learn, but we do it without being able to put all aspects of this knowledge into words. To use the above-mentioned example, we might be able to describe – in broad terms – what one should think about when learning how to ride a bike. But these instructions will not be enough to master the complicated motoric patterns necessary to execute the ability. This form of knowledge and learning-process instead demands practice. Procedural memory relies on the complex and interconnected performance of perceptual and motor pathways, involving, for example, the basal ganglia, neocortex, cerebellum, striatum, and the premotor- and primary motor cortex (see, e.g., Kandel et al. 2013). Many animals are endowed with procedural memory and are capable of procedural knowledge-how (Tulving 2002).

Semantic memory governs 'an individual's store of knowledge about the world. The content of semantic memory is abstracted from actual experience and is therefore said to be conceptual, that is, generalized and without reference to any specific experience.' (Binder and Desai 2011, p. 527). Semantic memory is crucial for numerous animals navigating a complex world (Roberts 2016). Moreover, an agent's ability to contemplate concepts and their relations, to perform inductive inferences and, as I want to emphasize, to *categorize* are all linked to semantic memory:

> Categorization is fundamental to understanding and using the concepts in semantic memory, since this process helps organize our knowledge and relate a test object to other known objects in the world. Categorization also allows us to engage in activities such as understanding unfamiliar objects and learning about novel objects. (Grossman et al. 2002b, p. 1549)

Gärdenfors and Stephens (2018) map conceptual knowledge-what (the knowledge of *what* a category consists in: dogs *characteristically* have four legs) to semantic memory. Similar formulations are indeed already in use in neuroscientific discussions:

> Thus humans use conceptual knowledge for much more than merely interacting with objects. All of human culture, including science, literature, social institutions, religion, and art, is constructed from conceptual knowledge. We do not reason, plan the future or remember the past without conceptual content – all of these activities depend on activation of concepts stored in semantic memory. (Binder and Desai 2011, p. 527)

Furthermore, several fMRI studies link the neural correlations of semantic encoding and semantic processing to '[...] many cognitive tasks, from perception, categorization, to explicit reasoning in problem-solving and decision-making.' (Goel and Dolan 2000, p. 110). In fact, many findings directly link semantic memory and categorization. Although, some discrepancies in neural activation is to be expected depending on, among other factors, variability regarding which aspects is in focus and regarding how stimulus is presented to test subjects (see, e.g., Grossman et al. 2002a). For example, Yee et al. (2014) explicitly relate conceptual knowledge to semantic memory and claim that such knowledge is distributed over many brain regions, which makes it flexible and able to handle varying contexts.

Semantic memory relies on associative pathways, involving, amongst other areas, the prefrontal cortex, the lateral-, ventral- and medial temporal cortex, basal ganglia, and hippocampus (see, e.g., Kandel et al. 2013):

> Semantic knowledge is stored in distinct association cortices and retrieval depends on the prefrontal cortex. [. . .] Semantic knowledge is distinguished from episodic knowledge in that it is typically not associated with the context in which the information was acquired. It is stored in a distributed manner in the neocortex, including the lateral and ventral temporal lobes. (Kandel et al. 2013, pp. 1449–1450)

Lastly, propositional knowledge-that (the knowledge *that* Stockholm is the capital of Sweden) maps to declarative episodic memory, governing factual remembrances and a sense of time – thereby playing a large part in how agents plan for the future:

> Memory for specific experiences is called episodic memory, although the content of episodic memory depends heavily on retrieval of conceptual knowledge. Remembering, for example, that one had coffee and eggs for breakfast requires retrieval of the concepts of coffee, eggs and breakfast. Episodic memory might be more properly seen as a particular kind of knowledge manipulation that creates spatial-temporal configurations of object and event concepts. (Binder and Desai 2011, p. 527)

Gärdenfors and Stephens (2018), ascribe facts to episodic memory (propositional knowledge-that) rather than to semantic memory (conceptual knowledge-what) – an interpretation somewhat similar to how for example Renoult et al. (2016) view ‘autobiographical facts’ as grounded in episodic memory. Episodic memory is crucially involved in self-awareness and first-person phenomenology. Since it according to Tulving’s account is an evolutionarily later specialized subcategory, as shown in Fig. 4.1, it is largely dependent on semantic memory:

> Episodic memory refers to a complex and multifaceted process which enables the retrieval of richly detailed evocative memories from the past. In contrast, semantic memory is conceptualized as the retrieval of general conceptual knowledge divested of a specific spatiotemporal context. [. . . T]he available evidence [. . .] converges to highlight the pivotal role of semantic memory in providing schemas and meaning whether one is engaged in autobiographical retrieval for the past, or indeed, is endeavoring to construct a plausible scenario of an event in the future. It therefore seems plausible to contend that semantic processing may underlie most, if not all, forms of episodic memory, irrespective of temporal condition. (Irish and Piguet 2013, p. 1)

Episodic memory relies on attentional pathways, involving, for example, the prefrontal cortex, and the ventral-fronto- and medial temporal cortex (see, e.g., Kandel et al. 2013).

Episodic memory is conventionally considered uniquely human although there is increasing evidence indicating that animals – primarily rats, corvids, and great apes – have some form of episodic memory. For example Panoz-Brown et al. (2016, p. 2821; see also Roberts 2016) argue that ‘[. . .] rats remember multiple unique events and the contexts in which these events occurred using episodic memory and support the view that rats may be used to model fundamental aspects of human cognition.’ Clayton et al. (2001, p. 1483) contend that ‘[. . .] jays form integrated memories for the location, content and time of caching. This memory capability fulfills Tulving’s behavioural criteria for episodic memory and is thus termed "episodic-like".' Rilling et al. (2007, p. 17149) describe how their '[...] results raise the possibility that the resting state of chimpanzees involves emotionally laden episodic memory retrieval and some level of mental self-projection, albeit in the absence of language and conceptual processing.' As a last example, Allen and Fortin (2013, p. 10380) even claim that '[...] core properties of episodic memory are present across mammals, as well as in a number of bird species.'

Tulving (2005) discusses the issue of episodic memory in animals and points out that:

> It depends partly on what one means by episodic memory, partly on the kinds of evidence one considers, and partly on how one interprets the evidence. When episodic memory is defined loosely as 'memory for (specific) past events,' then the standard commonsense answer is that of course animals have it. (Tulving 2005, p. 35)

However, Tulving highlights the importance of a less anthropomorphic perspective than this 'commonsense' understanding. Focusing on mental time travel, which is an essential aspect of episodic memory in humans and a distinguishing trait, he argues that:

> [...] only human beings possess "autonoetic" episodic memory and the ability to mentally travel into the past and into the future, and that in that sense they are unique. (Tulving 2005, p. 4)

The issue might be impossible to conclusively settle, since there are valid arguments for a variety of interpretations that ultimately hinge on how one *choose* to interpret the relevant terms, theories and evidence. Nevertheless, even if one accepts that animals other than humans can have episodic memories; it is to a significantly lesser degree. This fits with the view that episodic memory (propositional knowledge) is evolutionarily subsequent to the two other forms of LTM (Tulving 1985, 2002, 2005).

As previously mentioned, a way to increase the plausibility of the above-described knowledge-account is to investigate whether it is possible to integrate with models – from other sciences – on other levels of analysis. Since '[t]he neural basis of behavior cannot be properly characterized without first allowing for independent detailed study of the behavior itself [...]' (Krakauer et al. 2017, p. 488), the next section will explore the possibility of such integration by using Kornblith's (2002) analysis and account of knowledge from cognitive ethology (see also, e.g., Mitchell 2003; Cellucci 2017).

4.3 Cognitive Ethology: Evolution and Reliability

'The biological study of animal behavior, including its phenomenological, causal, ontogenetic, and evolutionary aspects, is a discipline known as ethology' (Anderson and Perona 2014, p. 18). Ethology investigates animal behavior concentrating on natural environmental settings. Moreover, there is an ongoing discussion if such behavior is intentional – and if so, to what degree (see, e.g., Allen and Bekoff 1995;

Wynne 2007; Shettleworth 2010). From an ultimate perspective, knowledge can be seen as the result of a phylogenetic and genotypic adaptive (functional) process, which shapes the cognitive faculties of agents (see, e.g., Plotkin 1993; Avital and Jablonka 2000):

> What is actually meant is that knowledge is a complex set of relationships between genes and past selection pressures, between genetically guided developmental pathways and the conditions under which development occurs, and between a part of the consequent phenotypic organization and specific features of environmental order. (Plotkin 1993, p. 228)

Our cognitive faculties are the result of evolutionary processes that has formed our sense organs and cognitive architecture. So, our evolutionarily molded cognitive faculties enable, as well as constrain, what we know (Plotkin 1993, p. 162).

In addition to these innate features, agents can acquire knowledge by *learning*, an ontogenetic aspect '[. . .] indicating processes by which the individual, thanks to phenotypic modifications, accommodates to novel circumstances in the course of its life.' (Serrelli and Rossi 2009, p. 18). In connection to ethology, implicit learning and implicit memory are central '[. . . involving] a wide variety of brain regions, most often cortical areas that support the specific perceptual, conceptual, or motor systems recruited to process a stimulus or perform a task.' (Kandel et al. 2013, p. 1459). Implicit learning splits into non-associative and associative learning, where non-associative learning includes responses to repeatedly encountered stimulus, in the form of habituation, where an agent's response diminishes by repeated exposure to a stimulus, and sensitization, where exposure strengthens a response. Associative learning involves how agents learn to link (associate) different stimuli to each other, in the form of conditioning by stimulus, response, and grasped relationships (see, e.g., Kandel et al. 2013).

Non-associative and associative learning thus match procedural respectively semantic memory, and, even though the focus is on particular brain systems rather than on implicit memory generally, for example Ullman (2016) argues that:

> Procedural memory involves a network of interconnected brain structures rooted in frontal/basal-ganglia circuits, including frontal premotor and related regions, particularly BA 6 and BA 44. [. . .] This circuitry underlies the implicit (nonconscious) learning and processing of a wide range of perceptual- motor and cognitive skills, tasks, and functions [. . .] including navigation, sequences, rules, and categories. (Ullman 2016, p. 956)

Illuminating the cognitive ethological position, Kornblith (2002, see also 1993) offers a fruitful discussion about 'fitness' and how animals that have knowledge about their changing environment better survive and thrive.[1] In a more traditional epistemological terminology, he points out that cognitive ethology provides:

> [. . .] a large literature on animal cognition, and [how] workers in this field typically speak of animals knowing a great many things. They see animal knowledge as a legitimate object of study, a phenomenon with a good deal of theoretical integrity to it. Knowledge, as it is portrayed in this literature, does causal and explanatory work. (Kornblith 2002, pp. 28–29)

[1] For a critique of Kornblith's position see for example Bermúdez (2006).

According to Kornblith's interpretation, cognitive ethology supports a reliabilist account of knowledge where knowledge should be seen as demanding reliably produced true beliefs (*RTB*)[2]:

> [. . .] I will argue that the kind of knowledge that philosophers have talked about all along just is the kind of knowledge that cognitive ethologists are currently studying. Knowledge explains the possibility of successful behavior in an environment, which in turn explains fitness. [. . . W]e must appeal to a capacity to recognize features of the environment, and thus the true beliefs that [. . . someone] acquire will be the product of a stable capacity for the production of true beliefs. The resulting true beliefs are not merely accidentally true; they are produced by a cognitive capacity that is attuned to its environment. In a word, the beliefs are reliably produced. The concept of knowledge which is of interest here thus requires reliably produced true belief. (Kornblith 2002, pp. 29–30, 57–58)

As reliabilism is generally coupled with externalist forms of justification such as *truth-connectivity* and *reliability* where an agent does not need to have cognitive access to her beliefs, it fits well with the description of the nonconscious non-associative and associative learning (see, e.g., Kandel et al. 2013; Ullman 2016).[3] An integration of the cognitive ethological reliabilist account and the cognitive neuroscientific account is accordingly possible by focusing on the two evolutionarily prior forms of memory and knowledge – procedural memory (procedural knowledge-how) and semantic memory (conceptual knowledge-what).

4.4 Cognitive Psychology: Intuition and Deliberation

Cognitive psychology investigates how human mental processes, including knowledge, are connected to behavior, using both bottom-up and top-down methods.

On a psychological level of analysis, implicit memory, implicit learning, and non-associative learning are all seen as being linked to procedural memory (procedural knowledge). In various forms of behaviorism these concepts have been investigated with a focus on reinforcement and punishment. However, in many theories, explicit memory and explicit learning take a central place, governing rule learning, awareness, and active remembrance of facts, being linked to episodic memory (propositional knowledge) (Kandel et al. 2013):

> [Explicit memory] is the deliberate or conscious retrieval of previous experiences as well as conscious recall of factual knowledge about people, places, and things. [. . .] Explicit memory is highly flexible; multiple pieces of information can be associated under different circumstances. (Kandel et al. 2013, p. 1446)

[2] Kornblith argues that cognitive ethology 'gives us the *only* viable account of what knowledge is.' (Kornblith 2002, p. 135, my italics). However, he does not motivate this restriction in a convincing way – pointed out by for example Kusch (2005) – and so this aspect of Kornblith's otherwise fruitful ideas will not be heeded here.

[3] Episodic memory (propositional knowledge) governing self-awareness and first-person phenomenology, on the other hand, is more naturally linked to internalism and forms of justification such as *rationality* and *cognitive access* (Tulving 2005; Alston 2005).

As an in-between, semantic memory (conceptual knowledge) is involved in both implicit and explicit memory, being linked to associative learning, pattern recognition, categorization, and prototype-matching. Regarding conceptual knowledge and categorization, for example Csibra and Gergely (2006) inquire into how teaching, and learning from teaching, should be viewed as a key adaptation for the transfer of knowledge between humans (see also Gärdenfors and Högberg 2017; Gergely et al. 2007). To facilitate social learning and teaching, they highlight how pedagogy offers a possibility to transfer generalizable knowledge, instead of just factual information, from a (active) teacher to a learner. Such generalizable knowledge does not only pertain to a specific situation but can be applied in many different contexts, which is essential for the ability to categorize. Csibra and Gergely (2009) develops their thoughts on generalizable knowledge:

> If I point at two aeroplanes and tell you that 'aeroplanes fly', what you learn is not restricted to the particular aeroplanes you see or to the present context, but will provide you generic knowledge about the kind of artefact these planes belong to that is generalizable to other members of the category and to variable contexts. Moreover, the transmission of such generic knowledge is not restricted to linguistic communication. If I show you by manual demonstration how to open a milk carton, what you will learn is how to open that kind of container (i.e. you acquire kind-generalizable knowledge from a single manifestation). In such cases, the observer does not need to rely on statistical procedures to extract the relevant information to be generalized because this is selectively manifested to her by the communicative demonstration. (Csibra and Gergely 2009, p. 148)

This type of generic generalizable knowledge, associated with categorization, seems reasonable to view as conceptual knowledge-what. Gärdenfors and Högberg point out that 'communicating concepts' is an evolutionarily prior form of teaching to 'explaining relationships between concepts' (Gärdenfors and Högberg 2017, pp. 193–195). According to Gärdenfors and Högberg, 'communicating concepts' at its core involves pattern-recognition, linking it to categorization and conceptual knowledge-what. 'Explaining relationships between concepts', on the other hand, involves teaching of facts and symbolic language making it more readily linked to propositional knowledge-that (Gärdenfors and Högberg 2017, pp. 193–195).

A well-established position in cognitive psychology is that of the *dual process framework* (see, e.g., Lizardo et al. 2016). Specifically the (default-interventionist) *dual process theory* has been prominent, which divides mental processing into one unconscious implicit and one conscious explicit reasoning system (see, e.g., Bago and De Neys 2017; Lizardo et al. 2016; Huberdeau et al. 2015; Sloman 1993, 2014; Evans and Stanovich 2013; Kahneman 2011; Rugg and Curran 2007):

- *System 1* operates automatically and quickly, with little or no effort and no sense of voluntary control. (Kahneman 2011, p. 20)
- *System 2* allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration. (Kahneman 2011, p. 21)

System 1 (or Type 1) can be described as intuitive and heuristic whereas System 2 (or Type 2) is deliberate and analytical, where the slow analytical process tries to inhibit the faster intuitive process.

There are a number of alternative theories arguing that cognition should be seen as consisting of a single process, as well as theories arguing for the possibility of parallel additional and/or more fine-grained systems (see, e.g., Bago and De Neys 2017; Rugg and Curran 2007). But I will follow Smith and DeCoster (2000, p. 110) who argue that 'numerous models of dual-processing modes can be integrated and interpreted in terms of the properties of two underlying memory systems and that this integration will lead to new insights and new predictions in several substantive areas of psychology.' (see also, e.g., Goel et al. 2000; Goel and Dolan 2003):

> The architecture that supports the interaction between systems has been hinted at in the cognitive neuroscience literature. Anatomically, the brain includes multiple parallel frontal corticobasal ganglia loops [. . .]. The interactions among these loops can be interpreted as a set of gating mechanisms [. . .]. My proposal is that one such loop is the intuitive loop, though it is best characterized as jointly intuitive and affective. Deliberation, in contrast, involves a more anterior prefrontal corticobasal ganglia loop. One critical function of deliberation is to serve to gate or at least modulate the intuitive–affective loop. (Sloman 2014, p. 75)

'System 1 is generally described as a form of universal cognition shared between humans and animals [. . . and] System 2 is believed to have evolved much more recently and is thought by most theorists to be uniquely human.' (Evans 2003, p. 454; see also Evans and Stanovich 2013, p. 225):

> Although rudimentary forms of higher order control can be observed in mammals and other animals [. . .], the controlled processing in which they can engage is very limited by comparison with humans, who have unique facilities for language and meta-representation as well as greatly enlarged frontal lobes [. . .]. We are in agreement that the facility for Type 2 thinking became uniquely developed in human beings, effectively forming a new mind [. . .], which coexists with an older mind based on instincts and associative learning and gives humans the distinctive forms of cognition that define the species [. . .]. (Evans and Stanovich 2013, p. 236)

System 1 is thus arguably compatible with the aforementioned *RTB*-account from cognitive ethology, and the two evolutionarily earlier memory forms (and knowledge forms) from cognitive neuroscience since '[t]he capabilities of System 1 include innate skills that we share with other animals.' (Kahneman 2011, p. 21):

> System 1 is old in evolutionary terms and shared with other animals: it comprises a set of autonomous subsystems that include both innate input modules and domain-specific knowledge acquired by a domain-general learning mechanism. System 2 is evolutionarily recent and distinctively human: it permits abstract reasoning and hypothetical thinking, but is constrained by working memory capacity and correlated with measures of general intelligence. (Evans 2003, p. 454)

In other words can System 1, on a cognitive psychological level of analysis, be mapped to procedural memory (procedural knowledge) and semantic memory (conceptual knowledge), on a cognitive neuroscientific level of analysis, and to *RTB*, on a cognitive ethological level of analysis. System 1 is thus most naturally linked to externalist justification – even though semantic memory (conceptual knowledge) can be seen as an 'in-between,' containing both externalist and internalist elements.

By adding episodic memory (propositional knowledge), on the neuroscientific level of analysis, System 2, on the cognitive psychological level of analysis, can be illuminated. System 2 is viewed as 'the conscious, reasoning self that has beliefs, makes choices, and decides what to think about and what to do' (Kahneman 2011, p. 21). Episodic memory, governs conscious and active reflection where we have cognitive access to our beliefs, on the neuroscientific level of analysis. It thus makes it possible to account for internalist justification and 'the subjective experience of agency, choice, and concentration' (Kahneman 2011, p. 21), on the cognitive psychological level of analysis, which is needed to fully explain human cognition. The three memory systems, on the neuroscientific level of analysis, can hence explain both System 1 and System 2 on the cognitive psychological level of analysis. In support of such integration for example Lizardo et al. (2016) explicitly connect 'know how' and non-declarative representation to System 1, whereas 'know that' and declarative representation is connected to System 2:

> [. . . M]emory is divided into two main types, most commonly referred to as "declarative" and "nondeclarative" memory. Declarative memory (Type II) consists of consciously accessible memories of facts, symbols, and events, while nondeclarative memory (Type I) consists of relatively less accessible procedural knowledge, habits, and dispositions. The two kinds of memory are sometimes distinguished as "knowing that" and "knowing-how" [. . .], or "explicit" and "implicit" memory [. . .]. (Lizardo et al. 2016, section 3.2)

The discussion points in the direction of compatibility and a plausible integration of the models on the presented three levels of analysis.

4.5 Conceptual Spaces: Knowledge-What and Categorization

The conceptual spaces framework has been presented and developed by Gärdenfors as a complementary alternative to the conventional symbolic and subconceptual forms of representation (Gärdenfors 1990, 2000, 2014). It postulates geometrical structures, where 'phenomenal' quality dimensions are grouped into domains. Observations of objects can then, in accordance with their properties, be positioned in a dimensional region. Properties can thereafter be compared in regard of their relation, where relative proximity represents degree of 'similarity.' To fruitfully analyze properties, categories, and their relations, Gärdenfors proposes two definitional criteria that provides spatial structure:

> *Criterion P*: A *natural property* is a convex region of a domain in a conceptual space, and;
> *Criterion C*: A *natural concept [or category]* is represented as a set of convex regions in a number of domains together with an assignment of salience weights to the domains and information about how the regions in different domains are correlated.

The convexity of criterion *P* is thought to capture that if two objects, exemplifying a property, are located in a particular domain, then objects positioned *between* those objects will also exemplify that same property. Criterion *C* highlights that natural concepts and categories are based on one *or more* domains – a distinction that is lost in the traditional language-focused approach. Conceptual spaces thus offer the ability to add and adjust dimensions in a domain, making it possible to elucidate how they are similar and/or connected. Furthermore, conceptual spaces make it possible to clarify and explain category-formation and learning. So, by utilizing conceptual spaces and criteria P and C it is thus possible to model categorizations and the knowledge linked to them; i.e. conceptual knowledge-what (see, e.g., Gärdenfors 2000, 2014; see also Douven et al. 2013; Decock et al. 2013).

Focusing on knowledge-what and categorization, Gärdenfors and Williams (2001) specifically address how categorizations efficiently can be modeled with conceptual spaces. Even though their focus is on artificial intelligence, the framework has the ability to clearly show prototypes, independent dimensions, and similarity. They point out that '[t]here is a wealth of psychological data supporting the existence of prototypes and their key role in categorization' (Gärdenfors and Williams 2001, p. 387):

> In summary the key findings from psychological studies of categorization are (i) similarity judgments play a fundamental role in categorization and they are context sensitive, (ii) the degree of similarity is judged with respect to a reference object/region such as a prototype, (iii) category membership can be graded (discrete membership, if and when it exists, is considered to be a special case), and (iv) the psychophysical relationship between the stimulus and the response depends on the underlying categorization. (Gärdenfors and Williams 2001, p. 387)

I regard it to ultimately be up to any theoretician to investigate those domains and quality dimensions that are found to be of interest. But reconnecting to the above discussion about our evolutionarily molded cognitive faculties; there are some innate, natural, domains and quality dimensions for humans, and '[o]ur quality dimensions are what they are because they have been selected to fit the surrounding world.' (Gärdenfors 2000, p. 82).[4] Taken together, this strongly indicates that conceptual spaces are apt for investigating categorization and conceptual knowledge-what – the knowledge of what a category characteristically consists in.

4.6 Concluding Remarks

An integration of the neuroscientifically grounded knowledge-account with accounts from cognitive ethology and cognitive psychology has been shown to be plausible. Procedural and semantic memory, on a neuroscientific level of analysis, match an ethological reliabilist account, as well as System 1 from the psychological dual process theory. By adding episodic memory, on the neuroscientific level of

[4] More or less similar domains and quality dimensions can also be found for other animals (see, e.g., Lorenz 1973; for an illuminating classic discussion see also Nagel 1974).

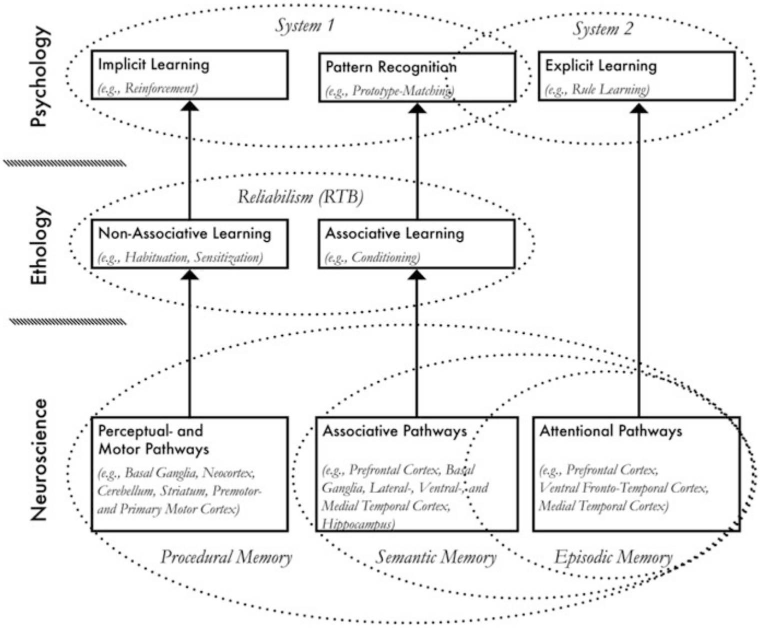

Psychology
System 1
System 2
Implicit Learning
(e.g., Reinforcement)
Pattern Recognition
(e.g., Prototype-Matching)
Explicit Learning
(e.g., Rule Learning)
Ethology
Reliabilism (RTB)
Non-Associative Learning
(e.g., Habituation, Sensitization)
Associative Learning
(e.g., Conditioning)
Neuroscience
Perceptual- and Motor Pathways
(e.g., Basal Ganglia, Neocortex, Cerebellum, Striatum, Premotor- and Primary Motor Cortex)
Associative Pathways
(e.g., Prefrontal Cortex, Basal Ganglia, Lateral-, Ventral-, and Medial Temporal Cortex, Hippocampus)
Attentional Pathways
(e.g., Prefrontal Cortex, Ventral Fronto-Temporal Cortex, Medial Temporal Cortex)
Procedural Memory
Semantic Memory
Episodic Memory

Fig. 4.2 Knowledge seen from a neuroscientific, ethological, and psychological level of analysis. Dotted lines indicate knowledge-categories; boxes outline examples of more detailed content descriptions, and; arrows show hierarchical mappings

analysis, System 2 on the psychological level can be accounted for. The article's integrative view is illustrated in Fig. 4.2.

This three-level naturalistic epistemological framework, linking conceptual knowledge-what to categorizations – fruitfully modeled within a conceptual spaces framework – promises interesting ramifications. On one hand it might fill a deleterious role and exert a dissolving influence on traditional epistemological problems and paradoxes. This is so since it moves a lot of focus away from propositional knowledge-that, which for a long time has had the center stage in epistemology, to conceptual knowledge-what. Moreover, it should impact discussions regarding, for example, reductionism since all three memory forms are considered important in their own right, which might be viewed as an argument against reduction. Importantly, *if* there is a reduction to be made it should be from propositional knowledge-that to conceptual knowledge-what and/or procedural knowledge-how, or from conceptual knowledge-what to procedural knowledge-how – *not* the other way around. On the other hand this nested take on naturalistic epistemology also offers a way to discover more and new scientifically grounded details regarding knowledge on other levels of analysis.

Acknowledgments I have had the great pleasure and privilege of investigating and discussing these topics with Peter Gärdenfors, and I am very grateful to Peter for sharing his vast knowledge, his eye for detail, and his positive energy. I would also like to thank Mauri Kaipainen for his generous and insightful remarks. Thanks to Trond Arild Tjøstheim for inspiring discussions. Finally I would like to thank my anonymous reviewers for comments.

References

Allen, C., & Bekoff, M. (1995). Cognitive ethology and the intentionality of animal behaviour. *Mind & Language, 10*(4), 313–328.

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences of the United States of America, 110*(Supplement 2), 10379–10386.

Alston, W. P. (2005). *Beyond "justification": Dimensions of epistemic evaluation*. Ithaca: Cornell University Press.

Anderson, D. J., & Perona, P. (2014). Toward a science of computational ethology. *Neuron, 84*(1), 18–31.

Avital, E., & Jablonka, E. (2000). *Animal traditions: Behavioural inheritance in evolution*. Cambridge: Cambridge University Press.

Bago, B., & De Neys, W. (2017). Fast logic? Examining the time course assumption of dual process theory. *Cognition, 158*, 90–109.

Barrett, H. C. (2015). *The shape of thought: How mental adaptations evolve*. Oxford: Oxford University Press.

Bermúdez, J. L. (2006). Knowledge, naturalism, and cognitive ethology: Kornblith's *Knowledge and its place in nature*. *Philosophical Studies, 127*(2), 299–316.

Berntson, G. G., & Cacioppo, J. T. (2009). Preface. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral science* (pp. xi–xii). New York: Wiley.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536.

Cellucci, C. (2017). *Rethinking knowledge: The heuristic view* (European studies in philosophy of science, Vol. 4). Springer.

Clayton, N. S., Griffiths, D. P., Emery, N. J., & Dickinson, A. (2001). Elements of episodic–like memory in animals. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 356*(1413), 1483–1491.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development. Attention and performance* (Vol. XXI, pp. 249–274). Oxford: Oxford University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

Decock, L., Dietz, R., & Douven, I. (2013). Modelling comparative concepts in conceptual spaces. In Y. Motomura, A. Butler, & D. Bekki (Eds.), *New frontiers in artificial intelligence* (pp. 69–86). Heidelberg: Springer.

Douven, I., Decock, L., Dietz, R., & Égré, P. (2013). Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic, 42*(1), 137–160.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge, MA: Harvard University Press.

Evans, J. S. B. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454–459.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.

Fantl, J. (2016). Knowledge how. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2016 Edition). https://plato.stanford.edu/archives/spr2016/entries/knowledge-how/

Fletcher, P. C., Büchel, C., Josephs, O., Friston, K., & Dolan, R. J. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex, 9*(2), 168–178.

Frank, M. J., & Badre, D. (2015). How cognitive theory guides neuroscience. *Cognition, 135*, 14–20.

Gärdenfors, P. (1990). Induction, conceptual spaces and AI. *Philosophy of Science, 57*(1), 78–95.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: Bradford Books/MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gärdenfors, P., & Högberg, A. (2017). The archaeology of teaching and the evolution of *Homo docens*. *Current Anthropology, 58*(2), 188–208.

Gärdenfors, P., & Stephens, A. (2018). Induction and knowledge-what. *European Journal for Philosophy of Science, 8*(3), 471–491.

Gärdenfors, P., & Williams, M. A. (2001). Reasoning about categories in conceptual spaces. In *Proceedings of the Fourteenth International Joint Conference of Artificial Intelligence* (pp. 385–392). Morgan Kaufmann Publishers.

Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science, 10*(1), 139–146.

Goel, V., & Dolan, R. J. (2000). Anatomical segregation of component processes in an inductive interference task. *Journal of Cognitive Neuroscience, 12*(1), 110–119.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*(1), B11–B22.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage, 12*(5), 504–514.

Goldman, A., & Beddor, B. (2016). Reliabilist epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/reliabilism/

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., & Gee, J. (2002a). The neural basis for category-specific knowledge: An fMRI study. *NeuroImage, 15*(4), 936–948.

Grossman, M., Smith, E. E., Koenig, P., Glosser, G., DeVita, C., Moore, P., & McMillan, C. (2002b). The neural basis for categorization in semantic memory. *NeuroImage, 17*(3), 1549–1561.

Horst, S. (2016). *Cognitive pluralism*. Cambridge: The MIT Press.

Huberdeau, D. M., Krakauer, J. W., & Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Current Opinion in Neurobiology, 33*, 71–77.

Irish, M., & Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in Behavioral Neuroscience, 7*(27), 1–11.

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus and Giroux.

Kan, I. P., Alexander, M. P., & Verfaellie, M. (2009). Contribution of prior semantic knowledge to new episodic learning in amnesia. *Journal of Cognitive Neuroscience, 21*(5), 938–944.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (Eds.). (2013). *Principles of neural science* (5th ed.). New York: McGraw-Hill, Health Professions Division.

Kim, H. (2016). Default network activation during episodic and semantic memory retrieval: A selective meta-analytic comparison. *Neuropsychologia, 80*, 35–46.

Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge: MIT Press.

Kornblith, H. (2002). *Knowledge and its place in nature*. Oxford: Oxford University Press.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron, 93*(3), 480–490.

Kusch, M. (2005). Beliefs, kinds and rules: A comment on Kornblith's *Knowledge and its place in nature*. *Philosophy and Phenomenological Research, 71*(2), 411–419.

Lizardo, O., Mowry, R., Sepulvado, B., Stoltz, D. S., Taylor, M. A., Van Ness, J., & Wood, M. (2016). What are dual process models? Implications for cultural analysis in sociology. *Sociological Theory, 34*(4), 287–310.

Lorenz, K. (1973/1977). *Behind the mirror: A search for a natural history of human knowledge*. London: Methuen.

Markie, P. (2013). Rationalism vs. empiricism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2013 Edition). http://plato.stanford.edu/archives/sum2013/entries/rationalism-empiricism/

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450.

Panoz-Brown, D., Corbin, H. E., Dalecki, S. J., Gentry, M., Brotheridge, S., Sluka, C. M., Wu, J., & Crystal, J. D. (2016). Rats remember items in context using episodic memory. *Current Biology, 26*(20), 2821–2826.

Pappas, G. (2017). Internalist vs. externalist conceptions of epistemic justification. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2017 Edition). https://plato.stanford.edu/archives/fall2017/entries/justep-intext/

Plotkin, H. C. (1993). *Darwin machines and the nature of knowledge*. Cambridge, MA: Harvard University Press.

Renoult, L., Tanguay, A., Beaudry, M., Tavakoli, P., Rabipour, S., Campbell, K., Moscovitch, M., Levine, B., & Davidson, P. S. (2016). Personal semantics: Is it distinct from episodic and semantic memory? An electrophysiological study of memory for autobiographical facts and repeated events in honor of Shlomo Bentin. *Neuropsychologia, 83*, 242–256.

Rilling, J. K., Barks, S. K., Parr, L. A., Preuss, T. M., Faber, T. L., Pagnoni, G., Bremer, D., & Votaw, J. R. (2007). A comparison of resting-state brain activity in humans and chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America, 104*(43), 17146–17151.

Roberts, W. A. (2016). Episodic memory: Rats master multiple memories. *Current Biology, 26*(20), R920–R922.

Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology, 7*(4), 532–547.

Rosch, E. (1975b). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*(6), 251–257.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Samet, J., & Zaitchik, D. (2014). Innateness and contemporary theories of cognition. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2014 Edition). http://plato.stanford.edu/archives/fall2014/entries/innateness-cognition/

Serrelli, E., & Rossi, F. M. (2009). *A conceptual taxonomy of adaptation in evolutionary biology*. Draft paper 4 september. Milano: University of Milano Bicocca.

Shettleworth, S. J. (2010). *Cognition, evolution, and behavior*. Oxford: Oxford University Press.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*(2), 231–280.

Sloman, S. A. (2014). Two systems of reasoning: An update. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 69–79). New York: Guilford Press.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108–131.

Stanley, J. (2011). *Know how*. Oxford: Oxford University Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1–12.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*(1), 1–25.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human. In H. Terrance & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 3–56). New York: Oxford University Press.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In G. Hickok & S. L. Small (Eds.), *The neurobiology of language* (pp. 953–968). London: Academic.

Wynne, C. D. L. (2007). What are animals? Why anthropomorphism is still not a scientific approach to behavior. *Comparative Cognition and Behavior Reviews, 2*, 125–135.

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2014). Semantic memory. In K. Ochsner & S. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience: Volume 1, core topics* (pp. 353–374). Oxford: Oxford University Press.

Paper IV

A. Stephens, T. A. Tjøstheim,
The cognitive philosophy of reflection
Erkenntnis, 2020 [2022], *87*, 2219-2242

Erkenntnis (2022) 87:2219–2242
https://doi.org/10.1007/s10670-020-00299-0

ORIGINAL RESEARCH

The Cognitive Philosophy of Reflection

Andreas Stephens[1] · Trond A. Tjøstheim[1]

Received: 31 January 2020 / Accepted: 21 July 2020 / Published online: 12 September 2020

Abstract

Hilary Kornblith argues that many traditional philosophical accounts involve problematic views of reflection (understood as second-order mental states). According to Kornblith, reflection does not add reliability, which makes it unfit to underlie a separate form of knowledge. We show that a broader understanding of reflection, encompassing Type 2 processes, working memory, and episodic long-term memory, can provide philosophy with elucidating input that a restricted view misses. We further argue that reflection in fact often does add reliability, through generalizability, flexibility, and creativity that is helpful in newly encountered situations, even if the restricted sense of both reflection and knowledge is accepted. And so, a division of knowledge into one reflexive (animal) form and one reflective form remains a plausible, and possibly fruitful, option.

1 Introduction

Throughout the history of Western philosophy, *reflection* has been considered an especially important human ability. Its role has long been prominent and can still be found at the center of theories by contemporary scholars such as, for example, BonJour (1985, 1998), Chisholm (1989), and Sosa (2007, 2009). Accordingly, a lot of effort has been invested in the inquiry of its role for thinking, knowledge, and justification. Common traditional positions have included that reflection is necessary in order to guarantee that an agent's knowledge is acceptable and certain, that her epistemic duty is fulfilled, that her knowledge is accessible, and that faulty beliefs due to inferential errors are avoided (see, e.g., Pappas 2017; see also Bortolotti 2011).

But in contrast to the above-described positions, Hilary Kornblith in his book *On reflection* (2012) points out that the common interpretation of reflection is problematic since reflection actually cannot provide that which many believe it can. Indeed much relevant research seems to indicate that rather than providing trustworthy

✉ Andreas Stephens
andreas.stephens@gmail.com

[1] Lund University, Box 192, 221 00 Lund, Sweden

knowledge, reflection can be quite unreliable. Numerous psychological studies, seemingly, show how human reflection often fails due to, for example, various biases (see, e.g., Stanovich and West 2000; Kahneman 2011). With this in mind, the importance of reflection, and its role for human thinking, knowledge, and justification, should arguably be deemphasized.

This leaves us at an interesting junction. On the one hand, reflection seems to underlie the very essence of human greatness and is commonly seen as a particularly important phenomenon. On the other hand, empirical evidence seems to support Kornblith's view and suggest that reflection only brings a false sense of certainty.

We recognize that inquiries are affected by the inquirer's stance (approach, commitments), which makes it important to briefly clarify our own. In line with Kornblith (see, e.g., 1993, 2002, 2012), we heed a naturalistic stance where philosophy needs to take relevant scientific results into account whenever such results are available. Accordingly, we accept both ontological and (cooperative) methodological naturalism, where natural phenomena and relevant scientific results are seen as more important than language or intuitions (see, e.g., Papineau 2016; Rysiew 2017; Cellucci 2017). We claim, as does Kornblith, that such a stance can offer philosophy new insights that are crucial for keeping the field relevant as well as for dissolving old problems.

In short, we believe that Kornblith's discussion of reflection is problematic due to its too-narrow understanding of what reflection brings to the table. Given this position, our aim in this article is to investigate reflection more broadly by examining relevant psychological constructs and their neural underpinnings. By stepwise investigating reflection on multiple levels of analysis, a synthesizing understanding of reflection that is biologically plausible can arguably be reached (see, e.g., Hassabis et al. 2017). This allows us to triangulate essential features of the natural phenomenon that Kornblith downplays or ignores (Horst 2016). We will, however, also argue that even if we accept a restricted view of reflection as 'second-order mental states,' as well as Kornblith's insistence on that reliability is the only epistemic value to consider, reflection, in fact, often does offer the subject added reliability. Importantly, this would leave the division of knowledge into a reflexive (animal) form and a reflective form a plausible option.

This article comprises five sections. In Sect. 2, we outline and discuss Kornblith's account of reflection. In Sect. 3, we investigate how reflection can be further elucidated by cognitive psychology, also outlining the neural correlates of reflection. In Sect. 4, we then explore philosophical consequences of the reached position pertaining to reliability and knowledge. Finally, in Sect. 5, we offer some concluding remarks.

2 Kornblith on Reflection

Kornblith (2012) argues that most traditional philosophers have valued reflection too highly due to faulty understandings of what it involves. And this overestimation has, in his view, led them to suggest, or even demand, that reflection is necessary when, in fact, such a view is wrong. Traditional philosophers, on Kornblith's view, tend to call on reflection when problems are recognized at a first-order level. Second-order reflection is then supposed to provide a solution by removing unreliability. This, however, according to Kornblith, is problematic since neither first-order processes nor second-order reflective scrutiny are entirely reliable. Kornblith argues that his points concerning reflection are generalizable and relevant for discussions of knowledge, reasoning, freedom of the will, and normativity. In this article we will focus on his discussion of knowledge.

Importantly, Kornblith addresses reflection specifically seen as consisting in 'second-order mental states.' He further considers reliability as being the only important criteria for belief acquisition processes (Kornblith 2012, p. 34). Kornblith then attacks the traditional view from two angles. Firstly, he argues that a reliance on reflection leads to an infinite regress and that reflection thus cannot provide the sought after reliability for first-order problems. Secondly, he argues that empirical evidence indeed indicates that the processes involved in reflection often are unreliable. Both these arguments, which will be presented more fully in the following subsections, according to Kornblith shows that reflection fails to be relevant for knowledge.

2.1 Infinite Regress

As a first argument against the traditional view, Kornblith claims that demands for reflection lead to an infinite regress since it continuously would require demands of ever higher-level reflections.[1]

According to Kornblith, knowledge, in its paradigmatic formulation, is commonly held to require justified true belief. And, as pointed out by Kornblith, according to many theoreticians, justification involves reflection on the epistemic status of one's beliefs. It is then only reflection that can guarantee the right epistemic status to one's beliefs. An omission to reflect would result in beliefs that cannot be considered knowledge.

We regard this a reasonable estimate of the common-sense view, although it arguably involves an implicit internalist view of knowledge. Indeed, Kornblith starts his discussion by presenting the famous 'Norman the clairvoyant' case by BonJour (1985). In short, BonJour (an internalist) argues that an agent needs active reflection, that makes her epistemically responsible, for knowledge. This is presented, by BonJour, as an argument against reliabilism (a form of externalism) that views knowledge as involving reliably produced true beliefs, hinging on the external connection between the agent and the world.

Now, Kornblith, who is an outspoken reliabilist (see, e.g., Kornblith 2002) argues that if an agent is to meet BonJour's requirements and reflect on her beliefs, the

[1] This same point plays out somewhat differently depending on which area of philosophy one is paying attention to, but we will, as aforementioned, here focus on knowledge.

reached beliefs would themselves, in turn, need to be justified by higher-order reflection, leading to an infinite regress (Kornblith 2012, pp. 12–13).

If one accepts Kornblith's strict understanding of reflection as second-order mental states and knowledge as being dependent on reliability, this indeed seems to be the forced conclusion.

2.2 Empirical Evidence Against the Reliability of Reflection

As a second argument against the traditional view, Kornblith claims that a wide range of empirical evidence shows that reflection often is unreliable. Reflective scrutiny does then most often not succeed in making us able to more reliably judge our first-order beliefs, but seems to make subjects more confident when in fact this is not motivated (Kornblith 2012, pp. 3, 25). This would indicate that it is not a tenable option to accept the aforementioned infinite regress as an inevitability and claim that having some reflective scrutiny at least is better than having none.

Sidestepping the merely logical matter of things, a large amount of empirical evidence seemingly does support Kornblith's interpretation where reflection is best seen as only bringing a false sense of certainty to the table. In defense of his position Kornblith presents, and interprets, several empirical findings that cohere with his account. Notably, he acknowledges the tentative nature of such findings and theorizing (Kornblith 2012, p. 136). It is also important to point out that Kornblith does *not* claim that reflection is useless, rather he argues that reflection might be useful if a more realistic account of it is accepted.

Kornblith focuses on cognitive psychology and the influential dual process theory. Briefly put, reflection figures distinctly in this framework, which partitions the mental into two forms. The first form (the old mind, System 1, or Type 1) is considered to be intuitive, automatic, non-conscious, and implicit, whereas the second form (the new mind, System 2, or Type 2) is reflective, controlled, conscious, and explicit.[2] On this account, the first form generate fast reflexive responses, which the second form sometimes reflectively inhibits (Tversky and Kahneman 1974, 1983; Sloman 1996; Barrett et al. 2004; Kahneman 2011; Evans 2007, 2008; Samuels 2009; Lizardo et al. 2016; Bago and De Neys 2017).

We consider Kornblith's choice to focus on dual process theory reasonable since that framework is canonical and directly addresses aspects of cognition that are highly relevant for understanding reflection and knowledge, being supported '… by a wide range of converging experimental, psychometric, and neuroscientific methods' (Evans and Stanovich 2013, p. 224). But, we want to point out that many interpretations of dual process theory exist, addressing, for example, types, systems or modes. This said, most interpretations of dual process theory can, arguably, be integrated into a common format which makes it fruitful to explore dual process theory as a, more or less, unified field although this should be done with care (Smith and DeCoster 2000, p. 110; Evans 2003, p. 458). Moreover, it should be mentioned that

[2] Kornblith uses the terminology 'System 1' and 'System 2' whereas, for example, Evans and Stanovich (2013, p. 226) argue against such a usage to the benefit of the 'Type 1' and 'Type 2' nomenclature.

there are researchers critical of dual process theory, where critics have pointed out both faults and alternative interpretations (see, e.g., Gigerenzer and Regier 1996; Keren and Schul 2009; Kruglanski et al. 2003; Osman 2004; Kruglanski and Gigerenzer 2011). The force of these lines of critique, though, hinge on which specific form of dual process theory they attack, and, for example, Evans and Stanovich (2013) in our view convincingly counters a number of the more common ones.

Importantly, if dual process theory, more generally, is not accepted as a provider of valid empirical input, Kornblith's argument would indeed be severely stifled. However, our main point here does not involve questioning dual process theory per se. Rather we claim that Kornblith's interpretation of cognitive psychological theorizing and evidence is problematic since it too narrowly *only* focuses on dual process theory. To remain a plausible option, Kornblith's restricted position needs to be developed in a pluralist direction that investigates the many important roles reflection fills for how a subject (organism) acts in her (its) environment (see, e.g., Shah and Vavova 2014). We will in the following Sect. 3 explore what such an account of reflection involves and how it can offer philosophy elucidating input.

2.3 Reflection as Decoupled from Knowledge

Taken together, Kornblith's arguments, indeed, seem to capture essential problems with the traditional positions that he criticises; it is, it seems, deeply questionable whether reflection can solve the problems often assumed that it can. And since reflection, indeed, does take such a center stage in much philosophical discussion, Kornblith's focus is highly relevant. Kornblith interprets the reached position as indicating that theoreticians ought to abandon any false hopes regarding what reflection can provide (Kornblith 2012, p. 7).

Kornblith discusses how Sosa's (1991; see also 2007; 2009) distinction between 'animal knowledge' and 'reflective knowledge' can offer a way out of the infinite regress. On this account, animal knowledge governs direct responses to one's sensory impacts, whereas reflective knowledge governs a wider understanding of one's responses and how they came about (Sosa 1991, p. 240). Animal knowledge is then more or less what externalist theories focus on, and reflective knowledge is what internalist theories focus on. Kornblith claims that this distinction, indeed, would resolve the issue of an infinite regress. Nonetheless he continues to argue that the reflective knowledge of the bisection does not add anything extra that is superior to 'mere' animal knowledge. Kornblith discusses, and rejects, the possibility that what reflective knowledge adds is increased reliability, which is also what Sosa argues (Kornblith 2012, pp. 16–17; Sosa 1991, p. 240). Since Kornblith considers reliability crucial for knowledge he then rejects a division of knowledge, even though he acknowledges that reflection might fill some other important role(s) (Kornblith 2012, pp. 19–20).

Yet, even if we accept the restricted view of reflection as second-order mental states, and accept that reliability is of sole importance (something we believe indicates a rather strong externalist position), then if it turned out that reflective processes do add to a subject's reliability, this would, on Kornblith's own account, rebut the infinite regress and make reflection eligible as underlying a distinct form of knowledge.

Kornblith accepts this possibility but emphatically denies that this is the case:

> We have examined a number of alternative motivations, and found that these motivations as well cannot bear the weight of the tempting distinction. It seems that there really is no ground at all for drawing a distinction between unreflective knowledge and something better, knowledge which involves reflection. (Kornblith 2012, p. 40)

We will in Sect. 4 specifically address how reflection *can* add reliability, even if the narrow account of it as only involving second-order mental states is accepted. This can be done by providing the subject with an opportunity to remember previous experiences and internally reflect on them in order to find patterns in them and then adjusting ensuing behaviors in accordance with the found patterns. In doing so the subject gains generalizability, flexibility, and creativity that is helpful in newly encountered situations. Therefore, a division of knowledge into one reflexive (animal) form and one reflective form remains a plausible, and possibly fruitful, option (see, e.g., Perrine 2014; Shah and Vavova 2014; Smithies 2016). So, although Kornblith (2012, pp. 16, 19) discusses how an allowance of two forms of knowledge could be seen as arbitrary and might risk leading to that infinitely many multiple forms must be allowed, we will below present a discussion that instead argues that two forms are biologically plausible.

But before we do this, we will next explore what a biologically plausible broader account of reflection involves and how it can offer philosophy elucidating input.

3 A Broader Understanding of Reflection

In this section, we follow Kornblith in focusing on cognitive psychology but, importantly, strive to stepwise develop a deeper multi-level investigation into reflection and its underlying processes that go beyond Kornblith's sole focus on dual process theory. This account, which also encompasses memory systems and neural correlates, offers a broader understanding of reflection that is not restricted to only involve second-order mental states. It is our belief that this account can provide philosophy with elucidating input that Kornblith's restricted focus misses.

In Fig. 1 we present a schematic illustration of how influential models from three levels of analysis cohere with each other, and how they relate to reflection. Although this is not an exhaustive account, we aim to substantiate this interdisciplinary approximation in the following discussion:

We now move to a description of how reflection is understood in cognitive psychology and find that a broader interpretation than the one Kornblith presents is motivated.

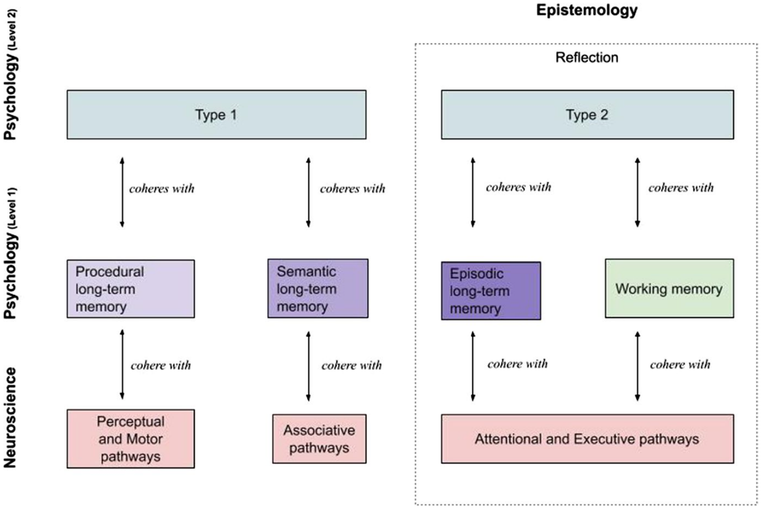


Fig. 1 Schematic illustration of relations between cognitive models, on different levels of analysis, and their relation to reflection. Four perspectives are represented: epistemology (dotted square); psychology level 2 (top row); psychology level 1 (middle row); neuroscience (bottom row). Boxes indicate model categories. Arrows indicate functional relationships

3.1 Reflection in Cognitive Psychology

In the dual process theory-literature, which is Kornblith's specific focus, reflection tends to be explicitly highlighted as an important phenomenon (see, e.g., Carruthers 2009; Mercier and Sperber 2009; Stanovich 2009; Evans and Stanovich 2013). According to dual process theory, reflection is considered to involve many specific functions linked to Type 2 processes (Evans 2008, p. 257). These complex functions encompass, for example, internal linguistics sequences or 'sentences of inner speech' (Frankish 2009, pp. 11–12; see also Carruthers 2009, p. 118), the ability to connect mental images to language, comprehend visual semantics, as well as visual manipulation (visual management) (Frankish 2010, p. 921; Carruthers 2009, p. 112). Moreover, from the perspective of dual process theory, the reflective mind is considered to include decision making, mental simulation, goal-adoption, belief-fixation, the ability for making comparisons, reasoning, metacognition in the form of second-order mental states, as well as hypothetical thinking (Evans and Stanovich 2013). Furthermore, recollection and the binding of information are dependent on reflection. It is crucial for a sense of time and to make out specific events (Yonelinas 2013, p. 2). In addition, Type 2 processes are linked to explicit rule learning (Evans 2008, pp. 257, 261, 267).

Even though human agents might not always be as in control as they believe themselves to be, these functions of reflection are important for their self-awareness and sense of agency. All these abilities are thus plausible to see as comprising a first outline.

There is a line of critique arguing that cognition is better seen as a continuum of processes than as two distinct ones (see, e.g., Osman 2004). This has some intuitive plausibility, however, by highlighting the difference of various *forms* of dual process theories this issue can, arguably, be circumvented. As Evans and Stanovich (2013, p. 229) point out, there are indeed *modes* of processing ('cognitive styles applied in Type 2 processing') that can vary on a continuum. Specific Type 2 reflections can thus be performed in a variety of different manners. But, what most dual process theories try to point out is that there are two distinct *types* of cognitive processes, where Type 2 processes stand out as being flexible and linked to reflection. And so, '[c]ontinuous variation in both cognitive ability and thinking dispositions can determine the probability that a response primed by Type 1 processing will be expressed—but the continuous variation in this probability in no way invalidates the discrete distinction between Type 1 and Type 2 processing' (Evans and Stanovich 2013, pp. 229–230).

So, even though there are pending issues concerning how we should view reflection from the perspective of cognitive psychology, we consider it initially plausible to link reflection to Type 2 processes. To reiterate, rather than viewing reflection as problematic, dual process theory indicates that it underlies several important cognitive functions such as internal linguistics sequences or 'inner speech,' visual semantic comprehension, visual manipulation, and mental simulation (visual management for short), decision making, goal-adoption, belief-fixation, reasoning, metacognition in the form of second-order mental states, hypothetical thinking, self-awareness, and our sense of agency.

To broaden our understanding of reflection and Type 2 processes we continue by focusing on a second, 'lower,' cognitive psychological level of analysis where the human memory systems are seen as consisting of many interconnected functional processes that encode, store, retrieve, and manage information. On this level, an influential division is made between long-term memory (LTM) and working memory (WM), where LTM can store information over a lifetime whereas WM governs active information handling (see, e.g., Repovš and Baddeley 2006).[3]

LTM is commonly partitioned into an implicit (non-declarative, non-conscious) system and an explicit (declarative, conscious) system. The non-declarative system is thought to govern automatic actions, whereas the declarative system is thought to govern abstracted knowledge about the world and autobiographical remembrance. In Tulving's (see, e.g., 1972, 1985, 2002, 2005) canonical and very influential three-part model of LTM, involving procedural, semantic and episodic memory, procedural memory governs perceptual and motor skills, semantic memory governs conceptual and categorical knowledge, whereas episodic memory governs remembrance of events (Tulving 1985, p. 2). According to Tulving '… procedural memory entails semantic memory as a *specialized* subcategory, and… semantic memory, in turn,

[3] This interpretation follows a development from previous traditional theories and models that placed a more passive short-term memory (STM) in the role now commonly ascribed to an active WM.

entails episodic memory as a specialized subcategory.' (Tulving 1985, pp. 2–3, italics in original).

Regarding WM, various models have been proposed although a very influential multi-component 'standard model' presents it as consisting of four parts: the phonological loop, the visuospatial sketchpad, the central executive, and the episodic buffer (Baddeley and Hitch 1974; Baddeley 2000, 2007; Repovš and Baddeley 2006; D'Esposito and Postle 2015; Chai et al. 2018). In short, the phonological loop controls auditory information, the visuospatial sketchpad controls visual and spatial information, the central executive controls attention and decisions, whereas the episodic buffer binds together information from different domains, working as a link to (episodic) LTM.[4]

Since it is through WM we actively handle information (see, e.g., Miller 1956; Cowan 2001) we argue that it is this system—on this level of analysis—which is primarily involved in Type 2 processes and reflection (Evans 2008). To substantiate this claim we show below how WM coheres with reflection as well as to the various previously mentioned features of Type 2 processes.

The phonological loop includes the articulatory network and the sensorimotor interface (Hickok and Poeppel 2007). It is thought to consist of a phonological store that can hold acoustic information for a couple of seconds, and an articulatory rehearsal process governing subvocalization by which verbal information is kept in memory. Apart from auditory information and speech, information needs to be re-coded through articulatory rehearsal before it can enter the phonological store. Accordingly, the phonological loop connects WM to language, and thus coheres with internal linguistics sequences and inner speech (Repovš and Baddeley 2006, p. 7).

The visuospatial sketchpad consists of two separate subsystems governing visual and spatial information respectively. It is crucially connected to how we perceive the world. Interestingly, we rely on a quite small amount of information from the surrounding world—since it tends to be stable, offering us a continuing 'external memory.' However, this bottom-up information also relies on top-down predictions when being interpreted into meaningful percepts (see, e.g., Friston 2010; Hohwy 2013; but see Firestone and Scholl 2016 for a recent challenge). The visuospatial sketchpad thus coheres with previously mentioned visual management abilities (Repovš and Baddeley 2006, pp. 8, 12).

The central executive is thought to be a form of control system for the other parts of WM (Rottschy et al. 2012, Sect. 1). By controlling attention, it governs how we

[4] There are alternative interpretations that, for example, argue that WM is best viewed as being a *part* of LTM (see, e.g., Ericsson and Kintsch 1995) or as an *emergent* property of numerous combinations of underlying 'possible subsystems' (see, e.g., Postle 2006), where '… working memory may simply be a property that emerges from a nervous system that is capable of representing many different kinds of information, and that is endowed with flexibly deployable attention.' (Postle 2006, p. 29). However, in line with for example Repovš and Baddeley (2006), we regard the empirical findings as providing a strong case for the standard model. Even so, we do acknowledge that it might have to be revised in a more fine-grained direction in light of coming findings, where feasible examples of such revisions might include, not only auditory- and visual-, but more subsystems based on all our different senses in WM.

prioritise, choose, and execute tasks. It is also involved in all information-manipulation (Repovš and Baddeley 2006, p. 14), composing reasoning as well as decision making and planning. But although being a central hub within WM, the central executive nonetheless has a limited degree of attention (see, e.g., Miller 1956; Cowan 2001). This means that the central executive coheres with abilities such as decision making, goal-adoption, belief-fixation, reasoning, metacognition in the form of second-order mental states, and hypothetical thinking.

The episodic buffer works as an interface between WM and LTM systems (Repovš and Baddeley 2006, p. 15). More specifically, it relates information between the central executive and episodic LTM '… forming a limited-capacity system for the ultra-short-term, intermediate storage of incoming sensory information' (Rottschy et al. 2012, Sect. 1). Through a store of limited capacity, it integrates information from the other components of WM into episodes. In doing so the episodic buffer is involved in creating conscious awareness. The episodic buffer binds recollected information, connecting to episodic LTM, which composes explicit rule learning (Strange et al. 2001, p. 1045). This interface thus processes and stores multi-dimensional representations (Rudner and Rönnberg 2008, p. 21). By doing so it helps to create a unitary experience, which is central for our self-awareness, sense of agency, and first-person phenomenological experience:

> Measures of working memory capacity have been shown to be predictive of performance in a wide variety of cognitive tasks… and highly correlated with fluid intelligence… It is the engagement of this system specifically that… has [been] emphasized in the definition of Type 2 processing and which underlies many of its typically observed correlates: that it is slow, sequential, and correlated with measures of general intelligence. [It] has also [been] suggested that Type 2 thinking enables uniquely human facilities, such as hypothetical thinking, mental simulation, and consequential decision making. (Evans and Stanovich 2013, p. 235)

In summary, we have shown that WM governs our internal linguistics sequences and connects to language (the phonological loop), our visual management (the visuospatial sketchpad), our attention, information-manipulation, reasoning, metacognition in the form of second-order mental states, and decision making (the central executive), as well as binds recollected information (the episodic buffer and episodic LTM). In view of the above discussion, we, therefore, claim that Type 2 processes and WM (also relying on episodic LTM) plausibly cohere with reflection.

3.2 Neural Correlates

By exploring the neural underpinnings of reflection, we in this subsection substantiate and ground our understanding of reflection in cognitive neuroscience. We argue that cognitive neuroscience is a suitable level at which to stop for our purposes, as this level provides information about plausible functionality of neural populations. Notably, such information can be effectively mapped to neural network architectures in a computer.

From the neuroscientific perspective, bottom-up perceptive pathways can be disassociated from top-down feedback pathways. The bottom-up pathways are activated by sensory stimuli, tending to align with statistical regularities in the sensorium by various process-signal amplifications (Pozo and Goda 2010). Collectively these processes contribute to the formation of distinctive receptive fields in the sensory cortices. The sensory streams are associated and bound together in association areas, which make up concept-like complexes that are presented to frontal populations involved in executive control (Tanaka 1996; Tsunoda et al. 2001; Caporale and Dan 2008; Magee and Johnston 1997; Ralph et al. 2010).

These frontal networks project back into the sensory pathways, which afford modulation of the perceptive streams via excitation and inhibition. This is the filter of attention, where certain aspects are turned down while others are amplified. Although the particulars of this process are still not fully known, there are indications that such top-down amplification is necessary to realize fine detail from a coarser bottom-up signal (Ahissar and Hochstein 2004).

Focusing on WM, it is closely associated with the processes and pathways of selective attention and executive control (Awh et al. 2006). Information may flow from the exterior world via the senses, or it may come from LTM.

The act of reflecting is, as described above concerning the phonological loop, often associated with internal linguistic sequences—internal monologues (Alderson-Day and Fernyhough 2015). An internal monologue involves both the production of speech as well as its interpretation. The former is realized by the posterior inferior temporal gyrus, premotor cortex, and the anterior insula, making up the articulatory network, along with the sensorimotor interface consisting of the sylvian parietal-temporal area (Hickok and Poeppel 2007). Interpretation, on the other hand, is realized by populations in the posterior middle temporal gyrus and posterior inferotemporal gyrus, making up the lexical interface (Kemmerer 2014). Semantic and grammatical aspects are integrated by the combinatorial network found predominantly in the lateral anterior temporal lobe. Together these pathways mediate understanding of conceptual content of speech. In short, this suggests that the articulatory network (posterior inferior temporal gyrus, premotor cortex, anterior insula), and the sensorimotor interface (sylvian parietal-temporal area) cohere with the phonological loop.

Although there are indications that all sensory modalities are available to WM (vision and audition: Baddeley and Hitch 1974; Baddeley 1986; tactility: Katus et al. 2012; proprioception: Smyth et al. 1988; olfaction: Zelano et al. 2009; somatosensation: Zhou and Fuster 1996), humans, as a species, are to a large degree reliant on vision in order to navigate and interact with the world (D'Ardenne et al. 2012; Brewer et al. 2011; Mason et al. 2007). The visuospatial sketchpad handles the visual and spatial information we encounter, which can be broken down into a number of sub-functions (Repovš and Baddeley 2006). For example, there appears to be a dissociation between purely visual representation, and representation of space as such (Constantinidis and Wang 2004). Spatial WM may be representing space generally, for visual, auditory, or other stimuli, and appears to be mediated by a network involving the dorsolateral prefrontal cortex, superior temporal cortex, posterior parietal cortex, and the lateral intraparietal lobe (Constantinidis and Wang 2004). These sites are lateral. On the medial side, the anterior cingulate cortex, posterior cingulate and retrosplenial cortices, and the parahippocampal cortex are involved (Constantinidis and Wang 2004). Parietal areas generally mediate integration of sensory streams, while the dorsolateral prefrontal cortex is usually thought to be responsible for maintaining and storing representations (though see Mackey et al. 2016 for a challenge to this in humans). Visual representations in particular also make use of networks in the occipital lobe (see, e.g., Schurgin 2018). These areas thus together cohere with the visuospatial sketchpad.

The most important cortical area for executive function, or cognitive control, appears to be the frontal cortex. A recent review by Badre and Nee (2018) identifies several regions within frontal cortices that mediate central executive control functionality of varying concreteness. In general, more abstract control is found in rostral areas, while concreteness increases caudally, closer to sensory cortices. Thus, the frontal eye fields and the premotor and motor cortices handle concrete sensory-motor control (Badre and Nee 2018). Contextual control is found more rostrally in the dorsal- and ventral anterior (pre) premotor areas, also including the inferior frontal junction area (Badre and Nee 2018). More rostrally still are areas that handle control of context-independent schemas. These include the mid-dorsolateral prefrontal cortex, and the rostrolateral prefrontal cortex (Badre and Nee 2018). In this context, schemas may be thought of as a kind of mental structures that organize classes of percepts and their relationships (Bartlett 1932). These, and other areas such as the frontostriatal circuits, brainstem, and superior parietal cortex cohere with the central executive.

As mentioned, the episodic buffer functions as a mediator between many memory systems, especially between the central executive and episodic LTM (Baddeley et al. 2010). When retrieval is needed for planning and executive control, the episodic buffer helps integrate relevant information (Strange et al. 2001, p. 1045; Rudner and Rönnberg 2008). Although the exact role and underpinnings of the episodic buffer remain unclear, particularly the parietal lobe and the left anterior hippocampus is thought to play a crucial role, in how this temporary storage, with a limited capacity, merge information (Berlingeri et al. 2008; Baddeley et al. 2010). This is enabled by a capacity for multi-dimensional coding, giving the episodic buffer a central role for conscious awareness, as well as for immediate- and episodic recall. Episodic memory is a broad concept, integrating sensory streams along with a sense of space, place, and time, but also a sense of agency. In the brain, this means that diverse and widespread networks are recruited to encode and reconstruct episodes. One of the most important networks is thought to be the hippocampus. Coarsely, it is responsible for spatiotemporal aspects of memory organization, as well as for relations between memories (Eichenbaum 2018). Also involved is the parahippocampal gyrus which more specifically processes aspects of place (Eichenbaum 2018). The ventromedial prefrontal cortex and the angular gyrus process self-referential aspects, and the feeling of agency respectively (Dede and Smith 2018). The middle temporal gyrus is thought to handle semantic aspects of episodes (Dede and Smith 2018). Included in episodic memory networks are neural populations related to attention. The retrosplenial and posterior cingulate cortices are involved in reducing attention and engaging the default network, which can reconstruct episodes. The ventrolateral prefrontal cortex is also thought to be able to break established attentional patterns to direct attention to other salient events (Corbetta and Shulman 2002; Eriksson et al. 2015). Similar mechanisms to manipulation of chunks may make up the affordance of mental time-travel and mental simulation, which appear to rely on recalling sequences from LTM and somehow parameterizing them. The hippocampus, in particular, appears to be involved with this, but likely in concert with prefrontal populations (Hassabis et al. 2007). Information from LTM route via the default network (Brewer et al. 2011; Mason et al. 2007). Specifically, there are indications that the fusiform gyrus, the inferior temporal and parahippocampal gyri, as well as the left posterior insula, are activated above baseline when gating of LTM is in effect (Brewer et al. 2011).

In this subsection, we have investigated the neural underpinnings of reflection and WM. Although the various parts of WM are interconnected, working in parallel with LTM and numerous other systems, a number of specific brain areas pertaining to selective attention and executive control do stand out. The articulatory network (posterior inferior temporal gyrus, premotor cortex, anterior insula), and the sensorimotor interface (sylvian parietal-temporal area) coheres with the phonological loop. The dorsolateral prefrontal cortex, superior temporal cortex, posterior parietal cortex, lateral intraparietal lobe, anterior cingulate cortex, posterior cingulate, retrosplenial cortices, and the parahippocampal cortex, as well as the occipital lobe, coheres with the visuospatial sketchpad. The frontal and prefrontal cortex, the premotor and motor cortices, also involving frontostriatal circuits, brainstem, and superior parietal cortex coheres with the central executive. The parietal lobe and the (left anterior) hippocampus coheres with the episodic buffer. And, the prefrontal, ventral fronto-temporal, medial temporal, retrosplenial, and posterior cingulate cortices, the parahippocampal, angular, middle temporal, the fusiform, and inferior temporal gyrus, as well as the left posterior insula and the hippocampus coheres with episodic LTM.[5] In short, the processes and pathways of selective attention and executive control cohere with WM and so Type 2 processes and reflection (Awh et al. 2006).

The reached position is thus that reflection involves Type 2 processes, WM and episodic LTM, as well as attentional and executive neural pathways. Reflection can

[5] Research on the cerebellum indicates that it plays a vital role not only in fine motor behaviour, but also in the automation of mental processes. According to Ito (2008), the cerebellum has two principal modi of operation: as a forward model, and as an inverse model. The former implies that the cerebellum can learn to generate and hence simulate sensory signals. The latter means that the cerebellum can learn to control, for example, muscles in the motor system, but may also be interpreted as to involve populations of excitatory and inhibitory neurons that affect contents of WM. Thus, the cerebellum can learn to perform volitional operations in WM automatically. Common examples of this is mental calculation, and certain kinds of planning (Ito 2008). This can be interpreted as the cerebellum being necessary for higher order thought, or being able to automate sequences of thought into building blocks that can be used for more complex problem solving or planning. Further aspects could, for example, include the function of glial cells in signal delay and the function of protein synthesis in regulating density of receptors or neurotransmitter reuptake mechanisms.

thus be differentiated from Type 1 processes, procedural and semantic LTM, as well as perceptual, motor, and associative neural pathways.[6]

We want to point out that even though this partitioning is well-established, highlighting an essential feature of human cognition, both reflexive and reflective processes involve complex intertwined bottom-up and top-down signals that work together. In the following Sect. 4, we will try to elaborate on this interaction.

3.3 Interpreting, Operationalizing and Measuring Reflection

Above, psychological constructs and their neural underpinnings, on multiple levels of analysis, have shown the natural phenomenon reflection to be multifaceted and complex, involving much more than just second-order mental states. This broader understanding of reflection thus provides input that more narrow accounts risk to miss. It is a dual understanding of cognition that emerges, which seemingly ought to influence our view of what a plausible account of knowledge should consist in.

But Kornblith questions the philosophical relevance of psychological findings and theories on the matter of reflection generally. He argues that there is an important difference between how ‘reflection’ is used in psychology and how it is used in philosophy (Kornblith 2012, pp. 141–142):

> While System 2 is often the source of second-order belief, not all of the beliefs produced by System 2 are second-order, and thus when psychologists speak of System 2 as involved in reflection, their use of that term better accords with everyday usage, which allows that we may reflect on various features of the world around us and not just on features of our mental life, than it does with the technical usage here which ties reflection to second-order states. (Kornblith 2012, p. 140)

Here Kornblith points out that he uses reflection in a technical sense. Accordingly, he accepts that Type 2 processes (System 2) involve other aspects, but considers that the only philosophically relevant aspect is the link to second-order mental states. From a cooperative methodological naturalistic perspective philosophers should look to science for answers rather than make up their own based on intuition, which makes it questionable to restrict scientific input in this manner. And as we have shown above, a broader interpretation is motivated. However, if the traditional view that Kornblith wants to counter demands that reflection is restricted to one of its aspects—second-order mental states—it might be necessary to do so for argument’s sake. It is then only the empirical evidence specifically addressing metacognitive second-order mental states that should be considered.

But Kornblith goes further. According to Kornblith, psychological theorists ‘mean to say nothing more [with the term reflection] than that the kind of thought characteristic of System 2 is conscious’ (Kornblith 2012, p. 141). Reflection

[6] Importantly, semantic memory is connected to both procedural and episodic memory although we will regard it as closer tied to reflexive generalized processes and thus not view it as directly involved in reflection (see, e.g., Binder and Desai 2011; Yee, Chrysikou, and Thompson-Schill 2014).

should then be understood as 'nothing more than' conscious reasoning in System 2 (Type 2 processes)—also involving non-conscious processes from System 1 (Type 1 processes). But we consider this interpretation to be insufficient and problematic. It is one thing to restrict one's focus (to second-order mental states)—against the scientific usage found in cognitive psychology. However, in claiming that cognitive psychologists (or even only dual process theorists) mean nothing more than 'consciousness' when they speak of reflection, we believe Kornblith is in the wrong.

Contrary to Kornblith's interpretation, cognitive psychologists point out how 'the reflective mind' governs our thinking dispositions, having a number of important specific roles, where 'reasoning and decision making sometimes requires both (a) an override of the default intuition and (b) its replacement by effective Type 2, reflective reasoning.' (Evans and Stanovich 2013, p. 236). Rather than indicating 'nothing more' than consciousness, reflection can be seen to encompass many particular states in human cognition, but importantly second-order mental states about one's own thoughts is a focal point where '[c]onclusions accepted for a reason are not intuitive but are, we will say, "reflective"… and the mental act of accepting a reflective conclusion through an examination of the reasons one has to do so is an act of reflection' (Mercier and Sperber 2009, p. 12).

Currently, a common way of operationalizing reflection in the context of cognitive psychology research is by means of the 'cognitive reflection test' (CRT) (see, e.g., Frederick 2005; Campitelli and Labollita 2010; Toplak, West, and Stanovich 2011; Vandekerckhove et al. 2014; Gronchi et al. 2016). The idea of this experimental test is to measure the disposition or ability of a subject to resist the first answer that comes to mind when posed with a set of questions. These questions are deliberately posed in a way to yield different answers if the subject uses quick intuitions, or if they deliberate and reflect. Here is a common example: *A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?*

The intuitive, quick answer is that the ball costs 10 cents. The correct answer, however, is 5 cents. The original CRT consists only of three questions, including the one posed above and two similar ones, and subjects are given the following instruction: *Below are several problems that vary in difficulty. Try to answer as many as you can*. The measure consists in counting the number of correct answers. Having said that, the test is usually not presented alone, but as part of a larger questionnaire where time and risk preferences are asked for. Perhaps unsurprisingly, studies using the CRT show a correlation between correct answers and reduced temporal discounting (Fredrick 2005). In other words, people that tend to answer correctly tend also to be more patient than those who go with the intuitive answer.

This is all very well, but what does it tell us about the epistemic value of reflection? First of all, it indicates that reflexive, or first-order beliefs may not always be reliable since there is a tendency for the brain to jump to conclusions when effort is involved in making an inference. Second, in the cases pertinent to the CRT, reflection is limited to second-order; i.e., there is no infinite regress. Thirdly, it implies that in many cases truth checking may have to be done with external support, e.g., with pen and paper. The point of this is only that representing symbols in the environment saves on mental energy as it were, since the symbols no longer have to be kept stable in the mind. This makes it less likely that energy saving processes get activated, which again can yield inaccurate conclusions.

In a sense, this can be interpreted as lending weight to Kornblith's criticism of reflection; it can be unreliable. However, importantly so can reflexive processes. The CRT supports that trains of thought can indeed be unreliable since the brain is prone to be miserly with its resources, and this can lead to inaccurate conclusions. But it appears that at least some of these limitations can be overcome by cognitive offloading onto the external world. Hence the process of second-order thought understood as truth checking intuitions can add reliability and epistemic value.

We have looked to cognitive psychology and gained a multi-level understanding of reflection going beyond second-order mental states, which has enabled a more informed interpretation. While this indicates the advantage of a broader understanding of reflection, we will in the next section grant the more restrictive view of reflection and knowledge. It will however be shown that even on such an account, a division of knowledge into a reflexive and a reflective form remains a plausible option.

4 The Plausibility of Two Forms of Knowledge

As shown in the previous section, reflection fills many important roles, but most crucially for our discussion we will in this section discuss how it adds reliability—even restrictively understood as 'second-order mental states,' which from a scientific perspective involves a view of reflection as consisting purley of metacognition. In accordance with Kornblith's own argument, a division of knowledge into one reflexive (animal) form and one reflective form thus remains a plausible option.

4.1 Reflection can Add Reliability

Reflection in fact does add reliability since a pure reliance on reflexive processes would in many cases be costly because observations risk being too context-specific (see, e.g., Smithies 2016). To test each encounter purely on the merits of current observational stimuli could even lead to disaster. The ability to run multiple test-scenarios, amounting to second-order mental states about previous trials, in one's head has great survival benefits. Agents can use reflection to generalize and abstract away non-essential information thereby gaining an overarching understanding and knowledge. A sole focus on reflexive processes thus risks to only allow specific context-dependent knowledge of specific cases. Reflection, seen as second-order mental processes (metacognition), adds generalizability, flexibility, and creativity that is helpful in newly encountered situations, and this, in turn, adds reliability (see, e.g., Olsson 2017a).

The bottom-up pathways that originate in sensory neurons can automatically associate with each other and with behaviour. By being exposed to a variety of stimuli, they can generalize in their own way and do limited extrapolations based on similarities, and on trial and error. These pathways have evolved to support survival and procreation, and are hence usually able to do an admirable job if left to their own devices. The limitation of the bottom-up pathways is in their context-specificity. If there is no outward similarity for the senses to latch on to, no behaviour will match. This can result in arbitrary and inappropriate behaviour, fearful behaviour and withdrawal from the situation, or anxiety and no behaviour at all. This is where top-down pathways, second-order mental states, and reflective behaviour comes in. Away from the situation, in a calm and safe place, sensory sequences can be recalled and be played back. Different alternative behaviours can be simulated and evaluated, amounting to thinking about one's thinking or second-order mentals tates, so as to hopefully cope better with similar situations in the future.

The top-down pathways, governing second-order mental states, can inhibit particularities in the sensory streams and hence discover common patterns in them. Particularities of instances of a category are often represented by higher frequency information, while commonalities tend to be represented by lower frequency information (Wiskott and Sejnowski 2002). In general, however, instance particularity is not limited to high frequencies, and full generalization requires an ability to inhibit any kind of property representation, be it shape, sound, or smell. Inhibition carries a burden of effort though (Dixon and Christoff 2012), and humans have learned to use external representations such as drawings to aid in abstract pattern identification and to reduce cognitive load (Risko and Gilbert 2016).

Reflection also affords the extraction of patterns from one context, and the re-concretization of those patterns into different contexts, using imagination to fill in required and appropriate detail. This can save a tremendous amount of energy that would otherwise be needed to arrive at the same behaviour in each specific context via trial and error. To be sure, large differences between the constructed scenario and the actual one may occur. And to an extent, the success of such an enterprise depends on the quality of the second-order models that are employed. That is, how well an agent understands the contexts in question. If both source- and target contexts are understood, re-concretization has a good chance of being successful, otherwise, the probability remains low. Even if the projected behaviour fails, a plan can still be made to gather information in the given context such that correct behaviour can be learned.

Crucially, during the reflective phase, information from cultural sources can be integrated to change behaviour. Human beings can communicate and exchange experience and knowledge, and through writing and reading that experience can be communicated across larger distances and over longer time spans. By means of writing, knowledge about the world can also accumulate over time affording later generations better cognitive methods and tools than previous ones. Such information integration is not possible purely by bottom-up experience of concrete situations, even if direct situational information is more accurate than that generated by means of reflection.

So, reflection, even if solely understood as second-order mental states (or meta-cognition), *can* add reliability through added flexibility and generalizability for the agent. In the next section, we will go into more depth about the contrast between reflective and reflexive knowledge from the perspective of feedback control.

4.2 Reflective and Reflexive Knowledge

Since it has been shown that reflection can add reliability, Kornblith's account can be evaluated anew. He agrees that if this is the case, the infinite regress (from Sect. 2.1) can be avoided. And this would leave the option of dividing knowledge into two forms, one reflexive (animal) and one reflective. In this subsection we elaborate on this possibility.

Even though the body (including the central nervous system with the brain) forms essentially a unified system under feedback control, it is nevertheless governed by distinct reflexive and reflective pathways (Pezzulo and Cisek 2016; see also, e.g., Friston 2009, 2010; Hohwy 2013). Top-down pathways continuously predict activity of bottom-up sensory pathways, while prediction errors make their way upwards in the hierarchy until they can be adjusted for by activating effectors. Here 'effector' is used as a broad term for processes that bind together and affect other processes, including, for example, low-level hormonal upregulation, reflexive motor actions initiated by spinal cord networks, as well as behaviour guided by high-level plans such as walking to a store to buy food, or even applying to college to get an education. So, albeit that human cognition and knowledge involve several complex intertwined capabilities, they are plausibly partitioned into a reflexive and a reflective form.[7]

Reflection can be interpreted as willful manipulations of WM content using such metaphorical effectors. This process can be applied to question and check the validity of spontaneous intuitions. Take the example from the CRT mentioned above, where the question is what the price of the baseball is given that both the bat and ball cost $1.10, and the bat costs $1 more than the ball. The spontaneous first-order thought is that the ball costs 10 cents. What reflection can do is to check more thoroughly if this is indeed the case. By laboriously setting up an algebraic equation and doing the math step by step, the original intuition can be scrutinized. In this case it was wrong; the mathematics yield the answer 5 cents. As long as this second-order process is trusted, as is usually the case with arithmetics, there is no need for further verification.[8]

Summing up, we claim that Kornblith is correct when he points out that traditional philosophical investigations often do not do justice to the natural phenomenon of reflection. Indeed, folk-psychological notions of reflection ought not to be allowed to take precedence or override scientifically grounded understandings of the natural phenomenon. But the reached conclusion is that philosophy needs to accept a pluralistic account of reflection and knowledge that acknowledges both reflexive and reflective processes that each provide specific information relevant for knowledge (see, e.g., Plotkin 1993; Alston 2005; Olsson 2017b). Moreover, Kornblith's

[7] This also holds true, to various degrees, for all mammals, and many other organisms (see, e.g., Allen and Fortin 2013; Carruthers 2013).

[8] Interestingly, the scientific process can be seen as an example of a kind of infinite regress, since there is seldom a 100% sure probability of experimental validity, and 100% validity can never in practice be reached. But experimental results can converge, which means that further experimentation becomes less urgent. Hence the regress, and the reflection, can be halted.

own interpretation of reflection is problematic, even given his own demarcations and demands. Importantly, there is a link between reflection and reliability making two forms of knowledge a plausible option—one reflexive (animal knowledge) and one reflective.

5 Concluding Remarks

We have shown that a better understanding of reflection is possible by looking at how it actually works. We have therefore moved away from a traditional stance focusing on language, concepts, certainty, and truth. Instead, we have adopted a naturalistic stance, in line with Kornblith, focusing on natural phenomena, scientific results, and plausibility. In accordance with this stance, we have explored how reflection coheres with the psychological constructs Type 2, WM, and episodic LTM, as well as to attentional and executive neural pathways. Importantly, reflection has been shown to fill a number of important functions: our inner dialogues, visual management, attention, information-manipulation, reasoning, decision making, metacognition, sense of agency, self-awareness, first-person phenomenology, remembrance, and awareness, motivating a pluralist account.

But we have also argued that this, more fine-grained, understanding of reflection, also acknowledging the influence and role of reflexive processes, does tie reflection to reliability by providing generalizability, flexibility, and creativity that is helpful in newly encountered situations. This indicates that the possibility to divide knowledge into a reflexive form and a reflective form is a plausible option, contrary to Kornblith's view.

Acknowledgements We want to thank Peter Gärdenfors, Martin L. Jönsson, Christian Balkenius, Maximilian Roszko, Asger Kirkeby-Hinrup, Erik J. Olsson, and Ingar Brinck for sharing their vast knowledge concerning this topic. We also want to thank participants at the Research Seminar in Theoretical Philosophy and the PhD Seminar in Philosophy at Lund University, and our anonymous reviewers for comments.

Funding Open access funding provided by Lund University. The authors gratefully acknowledges support from Makarna Ingeniör Lars Henrik Fornanders fond and Stiftelsen Elisabeth Rausings minnesfond: forskning.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*(10), 457–464.

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin, 141*(5), 931–965.

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences of the United States of America, 110*(Supplement 2), 10379–10386.

Alston, W. P. (2005). *Beyond "justification": Dimensions of epistemic evaluation*. Ithaca, New York: Cornell University Press.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139*(1), 201–208.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4*(11), 417–423.

Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.

Baddeley, A. D., Allen, R. J., & Hitch, G. (2010). Investigating the episodic buffer. *Psychologica Belgica, 50*(3), 223–243.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in Cognitive Sciences, 22*(2), 170–188.

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition, 158,* 90–109.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*(4), 553–573.

Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge: Cambridge University Press.

Berlingeri, M., Bottini, G., Basilico, S., Silani, G., Zanardi, G., Sberna, M., et al. (2008). Anatomy of the episodic buffer: A voxel-based morphometry study in patients with dementia. *Behavioural Neurology, 19*(1–2), 29–34.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536.

BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.

BonJour, L. (1998). *In defense of pure reason: A rationalist account of a priori justification*. London: Cambridge University Press.

Bortolotti, L. (2011). Does reflection lead to wise choices? *Philosophical Explorations, 14*(3), 297–313.

Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y. Y., Weber, J., & Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proceedings of the National Academy of Sciences, 108*(50), 20254–20259.

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making, 5*(3), 182–191.

Caporale, N., & Dan, Y. (2008). Spike timing-dependent plasticity: A Hebbian learning rule. *Annual Review of Neuroscience, 31,* 25–46.

Carruthers, P. (2009). An architecture for dual reasoning. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 109–127). Oxford: Oxford University Press.

Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences, 110*(Supplement 2), 10371–10378.

Cellucci, C. (2017). *Rethinking knowledge: The heuristic view* (Vol. 4). Dordrecht: Springer.

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology, 9,* 401.

Chisholm, R. M. (1989/1966). *Theory of knowledge* (3rd Ed.). Englewood Cliffs, NJ: Prentice Hall.

Constantinidis, C., & Wang, X. (2004). A neural circuit basis for spatial working memory. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 10*(6), 553–565.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201–215.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24*(1), 87–114.

D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences of the United States of America, 109*(49), 19900–19909.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology, 66*(1), 115–142.

Dede, A. J. O., & Smith, C. N. (2018). The functional and structural neuroanatomy of systems consolidation for autobiographical and semantic memory. *Current Topics in Behavioral Neurosciences, 37,* 119–150.

Dixon, M. L., & Christoff, K. (2012). The decision to engage cognitive control is driven by expected reward-value: Neural and behavioral evidence. *PLoS One, 7*(12), e51637.

Eichenbaum, H. (2018). What versus where: Non-spatial aspects of memory representation by the hippocampus. *Current Topics in Behavioral Neurosciences, 37,* 101–117.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211–245.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron, 88*(1), 33–46.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454–459.

Evans, J. S. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove: Psychology Press.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *The Behavioral and Brain Sciences, 39*(e229), 1–72.

Frankish, K. (2009). Systems and levels: Dual-system theories and the personal–subpersonal distinction. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 89–107). Oxford: Oxford University Press.

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass, 5*(10), 914–926.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives, 19*(4), 25–42.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences, 13*(7), 293–301.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? *Comment on Sloman. Psychological Bulletin, 119*(1), 23–26.

Gronchi, G., Righi, S., Parrini, G., Pierguidi, L., and Viggiano, M. P. (2016). Dual process theory of reasoning and recognition memory errors: Individual differences in a memory prose task. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 331–335).

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron, 95*(2), 245–258.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America, 104*(5), 1726–1731.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Horst, S. (2016). *Cognitive pluralism*. Cambridge, MA: MIT Press.

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience, 9*(4), 304–313.

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus and Giroux.

Katus, T., Andersen, S. K., & Müller, M. M. (2012). Nonspatial cueing of tactile STM causes shift of spatial attention. *Journal of Cognitive Neuroscience, 24*(7), 1596–1609.

Kemmerer, D. (2014). *Cognitive neuroscience of language*. New York: Psychology Press.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4*(6), 533–550.

Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge, MA: MIT Press.

Kornblith, H. (2002). *Knowledge and its place in nature*. Oxford: Oxford University Press.

Kornblith, H. (2012). *On reflection*. Oxford: Oxford University Press.

Kruglanski, A. W., Chun, W. Y., Erb, H. P., Pierro, A., Mannett, L., & Spiegel, S. (2003). A parametric unimodel of human judgment: Integrating dual-process frameworks in social cognition from a single-mode perspective. In J. P. Forgas, K. R. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 137–161). New York: Cambridge University Press.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review, 118*(1), 97–109.

Lizardo, O., Mowry, R., Sepulvado, B., Stoltz, D. S., Taylor, M. A., Van Ness, J., et al. (2016). What are dual process models? Implications for cultural analysis in sociology. *Sociological Theory, 34*(4), 287–310.

Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R., & Curtis, C. E. (2016). Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 36*(10), 2847–2856.

Magee, J. C., & Johnston, D. (1997). A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science, 275*(5297), 209–213.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science, 315*(5810), 393–395.

Mercier, H., & Sperber, D. (2009). Intuitive and reflective inferences. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 149–170). Oxford: Oxford University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.

Olsson, E. J. (2017a). Coherentism. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory* (pp. 310–322). London: Routledge.

Olsson, E. J. (2017b). Explicationist epistemology and epistemic pluralism. In A. Coliva & N. J. L. L. Pedersen (Eds.), *epistemic pluralism* (pp. 23–46). Cham: Palgrave Macmillan.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*(6), 988–1010.

Papineau, D. (2016). Naturalism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/naturalism/.

Pappas, G. (2017). Internalist vs. externalist conceptions of epistemic justification. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). https://plato.stanford.edu/archives/fall2017/entries/justep-intext/.

Perrine, T. (2014). Against Kornblith against reflective knowledge. *Logos & Episteme, 5*(3), 351–360.

Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences, 20*(6), 414–424.

Plotkin, H. C. (1993). *Darwin machines and the nature of knowledge*. Cambridge, MA: Harvard University Press.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139*(1), 23–38.

Pozo, K., & Goda, Y. (2010). Unraveling mechanisms of homeostatic synaptic plasticity. *Neuron, 66*(3), 337–351.

Ralph, M. A., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences of the United States of America, 107*(6), 2717–2722.

Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139*(1), 5–21.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., et al. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *Neuroimage, 60*(1), 830–846.

Rudner, M., & Rönnberg, J. (2008). The role of the episodic buffer in working memory for language processing. *Cognitive Processing, 9*(1), 19–28.

Rysiew, P. (2017). Naturalism in epistemology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition). https://plato.stanford.edu/archives/spr2017/entries/epistemology-naturalized/.

Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 129–146). Oxford: Oxford University Press.

Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attention, Perception, & Psychophysics, 80*(5), 1035–1056.

Shah, N., & Vavova, K. (2014). Review: On reflection by Hilary Kornblith. *Ethics, 124*(3), 632–636.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108–131.

Smithies, D. (2016). Reflection on: On reflection. *Analysis, 76*(1), 55–69.

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology Section A, 40*(3), 497–514.

Sosa, E. (1991). Knowledge and intellectual virtue. *Knowledge in perspective: Selected essays in epistemology* (pp. 225–244). Cambridge: Cambridge University Press.

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. I). New York: Oxford University Press.

Sosa, E. (2009). *Reflective knowledge: Apt belief and reflective knowledge* (Vol. II). New York: Oxford University Press.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic and autonomous minds: Is it time for a tri-process theory? In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645–665.

Strange, B. A., Henson, R. N. A., Friston, K. J., & Dolan, R. J. (2001). Anterior prefrontal cortex mediates rule learning in humans. *Cerebral Cortex, 11*(11), 1040–1046.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*(1), 109–139.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275–1289.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience, 4*(8), 832–838.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1–12.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*(1), 1–25.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human. In H. Terrance & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 3–56). New York: Oxford University Press.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315.

Vandekerckhove, M., Bulnes, L. C., & Panksepp, J. (2014). The emergence of primary anoetic consciousness in episodic memory. *Frontiers in Behavioral Neuroscience, 7,* 210.

Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation, 14*(4), 715–770.

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2014). Semantic memory. In K. Ochsner & S. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience: Volume 1, core topics* (Vol. 1, pp. 353–374). Oxford: Oxford University Press.

Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behavioural Brain Research, 254,* 34–44.

Zelano, C., Montag, J. M., Khan, R., & Sobel, N. (2009). A specialized odor memory buffer in primary olfactory cortex. *PLoS ONE, 4*(3), 829–839.

Zhou, Y., & Fuster, J. (1996). Mnemonic neuronal activity in somatosensory cortex. *Proceedings of the National Academy of Sciences of the United States of America, 93*(19), 10533–10537.

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

Paper V

A. Stephens, T. A. Tjøstheim, M. Roszko, E. J. Olsson,
A dynamical perspective on the generality problem
Acta Analytica, 2021, *36*(3), 409-422

Acta Analytica (2021) 36:409–422
https://doi.org/10.1007/s12136-020-00458-6

A Dynamical Perspective on the Generality Problem

Andreas Stephens[1] · Trond A. Tjøstheim[1] · Maximilian K. Roszko[1] · Erik J. Olsson[1]

Received: 21 July 2020 / Accepted: 11 December 2020 / Published online: 12 January 2021

Abstract
The generality problem is commonly considered to be a critical difficulty for reliabilism. In this paper, we present a dynamical perspective on the problem in the spirit of naturalized epistemology. According to this outlook, it is worth investigating how token belief-forming processes instantiate specific types in the biological agent's cognitive architecture (including other relevant embodied features) and background experience, consisting in the process of attractor-guided neural activation. While our discussion of the generality problem assigns "scientific types" to token processes, it represents a unified account in the sense that it incorporates contextual and common sense features emphasized by other authors.

Keywords Dynamical systems · Naturalistic epistemology · Reliabilism · The generality problem

1 Introduction

According to process reliabilism, justification of a belief amounts to being formed by a process that is reliably truth-conducive, and knowledge reduces to reliably produced true belief. An issue that has been raised against this theory is the generality problem, which is based on the observation that while a specific belief is always generated by a specific belief-forming process ("token" process), reliability only makes sense relative to repeatable processes (process "types"). Now since all tokens seem to belong to multiple types, each of which might differ in how reliable they are, it is unclear how it is determined which specific type a particular token belongs to and, consequently, how reliable the process is (see, e.g., Goldman 1979, 1986, 2017).

✉ Andreas Stephens
andreas.stephens@fil.lu.se

[1] Lund University, Lund, Sweden

To make the problem vivid, Conee and Feldman (1998) present an iconic example in which a person in an everyday situation looks out the window, sees a maple tree, and thus forms the belief that there is a maple tree outside. Conee and Feldman highlight how reliabilism needs to identify the type of the process causing the belief in the face of the fact that each token can be seen as belonging to multiple types:

> The token event sequence in our example of seeing the maple tree is an instance of the following types, among others: visually initiated belief-forming process, process of a retinal image of such-and-such specific characteristics leading to a belief that there is a maple tree nearby, process of relying on a leaf shape to form a tree-classifying judgment, perceptual process of classifying by species a tree located behind a solid obstruction, etc. The number of types is unlimited. They are as numerous as the properties had by the belief-forming process. Thus, process reliability theories confront the question of which type must be reliable for the resulting belief to be justified. It is clear that the answer to this question will significantly affect the implications of the theory. [...] So, which type has to be sufficiently reliable? (Conee and Feldman 1998, pp. 2–3)

As many authors have pointed out (Goldman 1986; Adler and Levin 2002; Comesaña 2006; Olsson 2016; Kampa 2018), the generality problem is not a problem just for process reliabilism but for many other epistemological theories facing similar issues. Thus, whether or not an epistemological theory is affected by the generality problem, or a similar difficulty, is unlikely to be an interesting concern when deciding among epistemological rivals. Even so, it is unsatisfactory that a fully convincing answer to the problem seems to be lacking.

We take as our starting point an observation made by Conee and Feldman themselves:

> The notion of reliability applies straightforwardly only to enduring mechanisms, such as an eye or a whole visual system, and to repeatable types of processes, such as the type: visually initiated belief formation. (Conee and Feldman 1998, p. 2)

It is, we will argue, indeed such a biologically centered perspective that stands the best chance of providing fruitful input to investigations of the generality problem.

Specifically, we will address two central concerns raised by Goldman and Beddor (2016; see also Olsson 2016; Goldman 2016):

> Which repeatable type should be selected for purposes of assigning a reliability number to the process token? If no (unique) type can be selected, what establishes the justificational status of the resulting belief? (Goldman and Beddor 2016, section 3)

In response, we will investigate the hypothesis:

> A token belief-forming process instantiates a uniquely "right type" of the biological agent's cognitive architecture (including other relevant embodied features)

and background experience, consisting in the process of attractor-guided neural activation.[1]

In section 2, we present our approach. In section 3, we discuss how dynamical features and complexity are important to take into account in order to reach a biologically plausible interpretation of belief-formation, justification, and knowledge. Section 4 contains a detailed example of our approach in action. In section 5, we present considerations that can be used to identify the type that a token instantiates in a particular case. In section 6, we mention and respond to a number of objections. In section 7, finally, we summarize our results.

2 Approaches to the Generality Problem

Conee and Feldman (1998, pp. 3–5) identify three conditions that an answer to the generality problem should satisfy. A reasonable account should be principled, plausible, and true to the spirit of reliabilism. The first condition rules out ad hoc solutions lacking a proper foundation. The second condition rules out type assignments that make implausible correlations between reliability and justification. The third condition excludes solutions that are not in line with the reliabilist epistemological tradition. Furthermore, according to Conee and Feldman (1998, p. 5), "[i]t is reasonable to look for a solution to the generality problem in three places: common sense, science, and context." According to the common sense approach, the relevant types are those that we use in our daily life: "seeing," "hearing" etc. The scientific approach consults a relevant science for guidance as to how to fix the types of given process tokens. The contextual approach holds that the relevant types are relative to context.

Concerning the common sense-approach, Conee and Feldman (1998, p. 7) claim that "there are far too many common sense types to provide a unique identification of the relevant type for each process token [and] not all beliefs resulting from any one such type are even approximately equally justified." However, Jönsson (2013) and Olsson (2016) have convincingly undermined these claims by providing strong arguments showing that people do converge (in line with basic level effects) on the same type description when they report their classification of belief-forming processes and justifiedness—contrary to Conee and Feldman's claims (see, e.g., Rosch 1973; Webb and Graziano 2015; but see Jönsson 2015).

The approach we will heed instead focuses on scientific types. In particular, we will focus on scientific explanations based on cognitive science of how particular process tokens instantiate certain process types (see, e.g., Gigerenzer 1991; Goldstein and Gigerenzer 2002; Lee 2007; Rysiew 2017). Belief-formation processes take place in the natural world, and empirically guided scientific theories provide our best understanding of natural phenomena. Moreover, the part of the natural world in which belief-formation takes place is the brain of a cognitive agent. Relying on the best scientific accounts currently on offer is very much in the spirit of naturalized epistemology (Quine 1969), which is the framework within which most reliabilists arguably situate

[1] Both process tokens and process types should be understood as involving mostly automatic and non-conscious processes, an interpretation that will be motivated in the following sections.

their theory. Importantly, we consider scientific types to involve contextual features, and we think that they can be used to understand common sense types. We hence consider it possible to unify the approaches highlighted by Conee and Feldman. We will return to this topic below where we will discuss how contextual factors influence how process tokens actually instantiates process types.

We are not the first to suggest a cognitively informed perspective on the generality problem: Alston (1995)[2] is another case in point. While we agree with Alston's observation that process tokens belong to specific natural kinds, we part company with him in that we do not consider conscious psychological processes to be the kind of processes we ought to focus on (see also Goldman 1979, 1986; Heller 1995):

> If the epistemic status of a belief is a function of the reliability of the process that generates the belief, it is the reliability of the *psychological* process that is crucial. Looking at perceptual belief formation, no matter how exemplary and no matter how finely tuned the neural transformations involved in the pathway from the eye to the brain, if the belief is not formed on the basis of the conscious presentation (and/or its neural correlate) in a truth-conducive way, the belief will lack the epistemic desideratum that is stressed by reliabilism. (Alston 1995, p. 12, italics in original)

Since reliabilism is an externalist theory of justification and knowledge, Alston's demand for internalist transparency seems to be misplaced (cf. Comesaña 2006, pp. 30–31). The important aspect for the externalist account is, rather, the truth-connection—whether a belief is formed in a way that reliably connects it to the world (Goldman and Beddor 2016).

As we will try to make plausible, particular process tokens automatically and non-consciously instantiate particular process types. It follows that an agent might not from her subjective first-person perspective be able to identify the relevant type in a given case. Moreover, the cognitive underpinnings of the process type will involve much more complexity than tends to be acknowledged even by authors with a cognitive bent. For these reasons, approaches that take the transparency of an agent's justification and knowledge for granted, demanding a linguistic rationalization concerning the correct type, are, in our view, biologically implausible.

Another naturalistic perspective on the generality problem is offered by Beebe (2004). He presents a two-step solution where he in the first step argues that a specific set of conditions must be satisfied by the relevant type. "According to the tri-level condition, cognitive process types are information-processing types that are partially defined by their computational and algorithmic properties" (Beebe 2004, p. 180). Here, Beebe follows Marr (1982) who presents a three-level hypothesis of cognitive processing (computational, algorithmic, and implementational). We consider this approach interesting but part way with Beebe (2004, p. 183) when he explicitly argues against the relevance of the implementational level claiming that "[a]lthough physical properties make important contributions to scientific explanations, they cannot help in selecting relevant cognitive process types." We consider this move to be problematic and will

[2] Alston focuses on what he calls psychological functions, habits, dispositions, or mechanisms. For a defense, see Adler and Levin (2002). See Feldman and Conee (2002) and Comesaña (2006) for criticism.

below argue that the implementational level does offer fruitful input to the generality problem. In fact, as Beebe points out, whereas the computational and algorithmic levels are abstract, the implementational level directly addresses natural physical phenomena in the world. But, far from being problematic or irrelevant, we will show below that these physiological underpinnings of belief-formation do elucidate pertinent information concerning how token belief-forming processes can be assigned a uniquely right type. The tri-level condition step is, however, only seen as providing a first delimitation of relevant types and we will not here address the second step of Beebe's solution concerning statistical relevance.

We will now proceed to flesh out how process tokens instantiate process types, using an approach that centers on scientific explanations of the natural phenomena involved in belief-formations at an implementational level.

3 A Dynamical Perspective on Process Types

Cognitively speaking, a specific belief is always generated by a specific belief-forming process (a token), consisting in the activation of a particular neural pattern. A particular process token (of neural pattern activation) might intuitively seem to be assignable to different process types of varying reliability, suggesting that a cognitive or dynamical approach to the generality problem is a non-starter. However, we nowadays have significant theoretical insight and much relevant empirical evidence concerning human neural activity showing that belief-forming processes follow specific paths—*attractors* (which we will present and discuss below) (Buzsáki 2006; Lakoff and Johnson 1999; Kinzler and Spelke 2007).

Importantly, while there is a language-focused sense in which tokens hypothetically can be said to belong to multiple types, there is, in fact, an "explicit rule," to use Goldman's (2016) expression, operating in the cognitive domain: a given belief-forming token process instantiates a unique process type of attractor-guided neural activation. In this section, we will support this claim by shedding further light on the cognitive reality of belief-formation.

Since the occurrence of attractor-guided neural activation is ultimately an empirical question, we will now discuss how beliefs actually get formed, which is best elucidated by a pluralistic approach (Dale et al., 2009). As an organism interacts with its environment, statistical regularities will be learned and associated with suitable actions. This process can continue throughout an organism's lifetime, but the rate of learning tends to decrease with time. When stimuli, such as a physical object, are encountered by the organism, it will activate different pathways with an intensity depending on how well the stimuli fits the ideal or prototypical pattern associated with the different pathways. Although several pathways may be activated at the lower levels of perception, winner-takes-all mechanisms are in place to increase the probability that only a single process is engaged at the highest level. Furthermore, expectations from experience tend also to favor a single outcome by biasing some pathways over others (see, e.g., Ward 2002; Buzsáki 2006).

In order to address the generality problem, we will use a dynamical perspective focusing on dynamical systems, or systems that change with time, such as the brain. When constructing a dynamical model, one of the first tasks is selecting a set of *state* *variables* that describe how the state of the system progresses in time. Each state variable is a dimension in the system's *state space*. In a model of the nervous system, the state variables might include the activity of various brain regions, while in a model of a biological cell, the variables might be the concentration of various molecules. How the system as a whole evolves through time is called its *trajectory*, and is dependent on how the system is *parameterized*. In contrast to variables, parameters are static for a given trajectory. The complete set of trajectories corresponding to all possible settings of the system's parameters is called the *flow* of the system. By observing the flow of a system, it is possible to identify patterns in the trajectories. Such patterns are usually regions or points in the state space where trajectories tend to end up. Since these regions appear to be attractive to nearby trajectories, they are often called *attractors*. So, processes do not follow random paths, but rather automatically move towards specific ones. Buzsáki (2006) illustrates a (limit cycle) attractor as follows:

> [T]hink of a racing car on a circular track. [...] The exact path of the car will vary somewhat in each round, bypassing other cars at different parts of the track, but this path variability is limited by the physical barriers of the track. The car can occupy any part of the track but should avoid the barriers. Thus, the track surface can be conceived of as the attractor of the car's orbit. (Buzsáki 2006, p. 137)

Using this metaphor, the car can be seen as a process token, whereas the track is the relevant process type. Every time the car (token) completes a lap, its tires affect the track (type), making the impressions deeper (the type more entrenched).

From a dynamical perspective, a discussion of the maple tree example will center on the human perceptual system, where primitive perceptual categories such as lines with varying orientation, and the complexes (such as branches, and later on whole maple trees) formed by these primitives further along the visual pathway, are attractors (scientific types) in the visual system's state space. A specific perceptual experience will then consist in a particular neural activation pattern that is the strongest, in accordance with the process of attractor-guided neural activation.

Now, we have argued that "the right" process type is to be identified with the process of attractor-guided neural activation, which clearly exemplifies Conee and Feldman's scientific types. However, an agent is affected by her context, and background experience, which is governed by, for example, evolutionary, developmental, social, and cultural factors. So, scientific types have clear elements of context-dependence built into them, in the sense that an agent's interplay with the external world affects which attractor is activated. Since most humans share a similar environment, the type that is instantiated (the process of attractor-guided neural activation) will—with small fluctuations depending on level of expertise—be the same for all agents. In section 5 below, we will reconnect to this topic at greater length. Given this, it is hardly surprising that agents report similar type assignments when prompted. Thus, it is in virtue of actual type-convergences (scientific types: attractors) concerning biological cognitive agents' mental and behavioral processes, including belief-forming processes, that our intuitive classifications (common sense types) tend to converge (Olsson 2016; Jönsson 2013). We thus believe that Conee and Feldman's three perspectives—common sense, science, and context—far from being mutually exclusive can be naturally combined in one unified account.

So, similar to how Beebe (2004) uses his first-step tri-level condition (on the computational and algorithmic levels) to demarcate relevant from irrelevant types, we have argued that the implementational level shows how there is a natural phenomenon underlying specific belief-formations amounting to a particular type—in Alston's parlance, a natural kind.

4 A Concrete Example

We here present a slightly more detailed account of the features involved in visual processing (seeing a maple tree), which is the scene that Conee and Feldman introduced. As we have claimed, a given belief-forming token process instantiates a unique process type of attractor-guided neural activation.

From two-dimensional information, agents are able to create a three-dimensional world. In short, segregating processes keep different objects separate, by means of, for example, visual processing of stimuli into foreground (object) and background. This is done by an interplay of bottom-up aspects and top-down aspects together forming a percept. Bottom-up aspects involve, for example, recognition of edges and other basic visual cues, whereas top-down aspects involve, for example, past learning and background experience. Grouping processes, on the other hand, let the agent assemble elements into wholes. Groupings thus enable Gestalts (meaningful wholes) to be perceived through factors such as proximity, similarity, common fate, continuation, and closure. An additional important factor that greatly affects what is perceived is the agent's attention, since the spotlight of attention governs what stimuli will be processed. Furthermore, the complete visual scene that makes up the agent's visual stimuli will, with its different cues, interact to form a contextual understanding in the agent (Kandel et al. 2013, pp. 611–615).

Another way to describe this whole process type is that low-level processings of orientation, color, contrast, disparity, and movement direction are translated into intermediate-level processings of color integration, surface properties, shape discrimination, surface depth, surface segmentation, and object motion, which then are integrated into a high-level identification, by way of tying together visual primitives with categorical and associative linkings, additional sensory signals, memories, emotional valence, and top-down predictions (Kandel et al. 2013, pp. 560, 622; see also Friston 2010; Clark 2013, 2015). Iterations of the visual perceptual process in general, as well as specific processes for particular stimuli, affect (form) the agent's cognitive faculties:

> [C]ognition is nothing more (and nothing less) than a special kind of pattern formation, the interplay of functional segregation and integration and the continual emergence of dynamical structures that are molded by connectivity and subtly modified by external input and internal state. The shape of cognition, the nature of the information that can be brought together and transformed, is determined by the architecture of brain networks. The flow of cognition is a result of transient and multiscale neural dynamics, of sequences of dynamic events that unfold across time. The variety of cognition, the seemingly endless diversity of mental

> states and subjective experiences, reflects the diversity and differentiation made possible by the complexity of the brain. (Sporns 2011, p. 206)

In Conee and Feldman's described scene, the agent centers her attention on a specific place outside her window. A specific phenomenon (a maple tree) is segregated out as a foreground object, in contrast to the background, and a maple tree Gestalt is recognized. The encountered stimuli are thus processed both bottom-up and top-down, together governing what is perceived. As the example is presented, light reaches the agent's eye. The agent's retinal photoreceptors (cones and rods) register and convert the incoming stimuli and send the encountered information forward via retinal ganglion cells and the optic nerves towards other brain areas, in particular, to the lateral geniculate nucleus (LGN). From here, information can be sent to the amygdala, if the agent has strong emotional connotations to maple trees. If so, an emotional reaction, as well as automatic behavior, might ensue. Regardless of emotional content, information is forwarded to the primary visual cortex (V1–V3) of the occipital lobe at the back of the skull, where various specialized neurons process the information. Different cell layers are sensitive to different aspects such as the color of the maple tree and its leaves (parvocellular) and the movement—if there is any—of the maple tree's branches (magnocellular). The information is re-coded and then follows two central neural streams: the ventral stream and the dorsal stream. The ventral stream focuses on visual identification of *what* the stimuli is (components of a maple tree). This stream goes off towards the temporal lobe. The dorsal stream focuses on *where* the stimuli is located (outside the window, standing vertically) and *how* an agent might interact with the stimuli (pluck leaves, chop down). This stream moves towards the parietal lobe at the top of the agent's skull. At the early stages of these streams, only basic featurescan be detect. By combining these features it is nonetheless possible for the agent to combine, at later stages, this information into more complex representations, in for example regions such as V4 and V5/MT.

Moreover, top-down processes originating from prefrontal areas at the front of the agent's skull connect with the bottom-up processes. These processes enable the agent to predict its incoming stimuli (see, e.g., Friston 2010; Clark 2013, 2015) together with memories linked particularly to temporal lobe areas and the hippocampi (Mizumori 2013; Smith and Bulkin 2014). This aspect thus elucidates a particular feature of how the context governs perception. To clarify, if the context makes it salient that the object is to be interpreted as a maple tree, the agent can do so even before processing all the bottom-up stimuli. This is so since rather than elaborating on all individual stimuli, the agent can save energy by only tending to relevant prediction-errors (Friston 2010; Clark 2013, 2015; Hohwy 2013). This means that if the agent is well-acquainted with the scene outside her window she will not have to expend a lot of energy to categorize what she sees. Instead, her background experience makes her have certain preconceptions about the relevant scene. As long as there is nothing that the agent interprets as being out of the ordinary in the scene, she will quickly categorize what she sees (as a maple tree).[3]

[3] If the tree, for example, had been cut down, the agent would—most likely—have interpreted this gap as a prediction-error.

What we intend to highlight with this outline is that the process type (the process of attractor-guided neural activation) involved in seeing a maple tree depends on very specific elements relevant for visual perception. Even though most brain functions involve a number of parallel processes, it is nevertheless the case that neural pathways for specific functions reside in specific brain areas. Indeed, these "[...] specific patterns of interconnection and the resulting functional organization of neural circuits in distinct brain regions underlie the individuation of behavior" (Kandel et al. 2013, p. 337). Hence, specific neural activations (tokens) do not take place in a black box. Instead, they follow specific pathways. And, even though all humans to a large extent overlap in how our cognitive architecture is shaped, all individuals have specific attractor pathways that are unique for them. Now, to reiterate, when someone sees a maple tree, a particular neural activation pattern will be the strongest, yielding a unique process of attractor-guided neural activation. This is the relevant type.

5 The Role of Context and Social Factors

Goldman and Beddor (2016, section 4) point out that it is generally assumed that "a 'solution' to [the generality] problem will consist in a formula for identifying a unique process type given any specified case and token (assuming the case is specified in reasonable detail)." As we have tried to make clear, this is a request that is possible to satisfy although doing so involves a high degree of complexity. In other words, the complexity of scientific types entails that, although there exists such a type in every concrete case, it might be difficult to describe or identify it. Furthermore, any particular specification can always be put in "theoretical doubt" by further demands concerning what amounts to reasonable detail.[4]

Above, we have identified how belief-forming process tokens instantiates a uniquely "right type" of the biological agent's cognitive architecture (including relevant embodied features) and background experience, consisting in the process of attractor-guided neural activation that can separate relevant types from irrelevant ones. Below, we elucidate a number of perspectives that delineates contextual and background experiential factors that influence how process tokens actually instantiates process types (Olsson 2016, pp. 180–181), where particularly hippocampal neuron ensembles play a crucial part regarding expectations and prediction-error detection of contextual features (see, e.g., Mizumori 2013; Smith and Bulkin 2014):

> When a new context is encountered, a unique hippocampal ensemble is recruited to represent it. Memories for events that occur in the context become associated with the hippocampal representation. Revisiting the context causes the hippocampal context code to be re-expressed and the relevant memories are primed. As a result, retrieval of appropriate memories is enhanced and interference from memories belonging to other contexts is minimized. (Smith and Bulkin 2014, p. 52)

[4] Although, for example, Peirce (1877) would argue against the relevance of this merely theoretical possibility.

As mentioned by Conee and Feldman, context is an important aspect to take into account regarding belief-formations:

> Although a solution must be principled, it need not state necessary and sufficient conditions for relevance that are either precise or always determinate. Claims to the effect that a belief is "epistemically justified" might be vague and they might be context-sensitive in various ways. A solution must be universal only in that it must specify the relevant type whenever there are definite facts about justification. (Conee and Feldman 1998, p. 4)

The situational context (including social factors) we find ourselves in is then what governs which combination of factors is made salient and deemed relevant, which in turn determines which process of attractor-guided neural activation is instantiated. While Conee and Feldman classify context as a perspective of its own, we instead see it as an aspect that enters into a scientific perspective (scientific types). For a given process token, the context helps determine a process type as the process of attractor-guided neural activation in that context (see, e.g., Alston 2005; Wunderlich 2003). For instance, depending on what amount of detail is required, different process types might be relevant. Although contextual differences exist, there are many more factors that remain across contexts. For example, evolutionary, developmental, social, and cultural factors are to a large extent universally present.

Concerning the development of human categorization abilities, Murphy (2002, p. 328) argues that human children develop the ability to distinguish basic level categories at around 2.5 years. This is followed by more abstract, superordinate categories at 4 years. The last to develop are more particular categories, or subordinate categories, which become available at about 5 years of age (Murphy 2002). Superordinate categorization intuitively implies being able to remove information from percepts, i.e., being able to do abstraction. However, according to Markman et al. (1980), it may be that children rather regard superordinate categories as collections of things. When considering that removal of detail in practice requires inhibition, a late-arriving ability in the history of evolution, the grouping hypothesis appears likely. The observation that subordinate categories take longer to develop than superordinate ones may also make sense from a perspective of statistics. Exemplars of subordinate categories are by nature sparse, hence it takes time for the child to have experienced a sufficient number of exemplars to arrive at the finer details. More generally, children tend to make use of *heterarchies* more often than proper hierarchies, and they appear to have a preference for a single level of classification (Murphy 2002, p. 327). In any case, the preference for basic level categories, as noted above, remains throughout adulthood (Murphy 2002).

Our social upbringing shapes our adult selves, involving for example parental and peer influence, as well as our position in dominance hierarchies. Moreover, social stereotypes tend to be used in order to arrive at fast judgments. Social animals have evolved the ability to efficiently understand each other through the development of body language and visual displays, scents, or noise patterns based on a similar perception of the world. Our biology thus limits the possibilities of divergent understandings of the world by facilitating similar interpretations of the world through the overlap of our DNA that code for the development of our sensory and communicative organs, and the structure of our interpretative brains. Humans are, as social beings, genetically predisposed to copy the type of higher-order cognitive processes other beings use to understand the world. Finally, cultural aspects such as education, norms, rituals, and "ways of lives" govern what we find "normal" and what we strive to accomplish.

From a dynamical and systems perspective (Mobus and Kalton 2015), natural phenomena can be investigated on different levels and from different perspectives. But assuming a naturalistic position, such investigations are mutually supportive rather than adversary. Importantly, it is not the role of philosophers, *qua* philosophers, to decide whether any particular science is correct or not. Hence, in response to theoreticians who think that the generality problem for reliabilism cannot be solved, we have argued that actual process tokens instantiates certain process types (attractors). Furthermore, we have pointed out a number of factors (evolutionary, developmental, social, and cultural) that separate relevant from irrelevant types and determine the existence of a unique type in a way that is in principle accessible to an investigator even though gaining such access would require a lot of work. However, if a solution to the generality problem is required to provide a procedure by means of which the unique process type in a given case can be easily identified, we believe that no such solution can be found—not because there are no unique types but because the demand vastly mischaracterizes the complexity of the world and our brains.

6 Objections

Conee and Feldman's arguments against the scientific approach center on Alston's (1995) account of natural kinds. As a first issue, they believe that "[m]erely citing the fact that each belief-forming process falls into a natural kind does not provide an adequate rule of relevance" (Conee and Feldman 1998, p. 10). They instead argue that "there is no good reason to think that each token belongs to just a single natural kind" (Conee and Feldman 1998, p. 10). As has been shown above, this is a mischaracterization of natural dynamical processes such as belief-forming processes. Process tokens do, in fact, belong to specific natural process types (attractors).

Since different sciences use different (terminological) categorizations, it is tempting to view them as indicating different natural kinds. Conee and Feldman use this point as basis for a second critique, although this also amounts to a mischaracterization. A certain natural phenomenon can indeed be *described* (investigated, modeled) by various sciences on different levels of abstraction. But this does not indicate that the natural phenomenon belongs to different natural kinds, rather a specific natural kind can be "triangulated" from many points of view (Horst 2016). Accordingly, a certain process of attractor-guided neural activation is an instantiation of a specific natural kind—although it can be described using different sciences.

Conee and Feldman continue by arguing against Alston's idea that a specific function of belief-forming process activation governs the relevant type. Above, we have mentioned, and argued against, Alston's claim that only internally salient types are relevant. But when this restriction is removed, Alston's account is basically correct—although it is not presented in a detailed manner. Conee and Feldman (1998, p. 11) mention how Alston's position entails that "there is only one actually operative 'psychologically real' type for each belief-forming process." and then critique Alston for being unclear regarding how the agent could "pick out" the correct type, out of all theoretically possible. We have argued that this is not something that the agent does consciously; rather, this is automatically done by non-conscious reflexive processes governed by the world.

Moreover, Conee and Feldman critique Alston for promoting functions that are maximally specific. It is then "only one function [which] is 'operative' in the formation of any belief" (Conee and Feldman 1998, p. 13). But this, according to Conee and Feldman (1998, p. 15), leaves reliabilist theories "unable to distinguish the epistemic status of lucky guesses that happen to be based on distinctive features from expert judgments based on well-understood classifications." As has been previously discussed, it is the world—not the agent—that is of importance regarding the governance of which type is relevant. This objection is therefore based on a faulty understanding of how tokens instantiates types.

Lastly, Conee and Feldman point out how "[t]here is no good reason to think that the types that are of greatest value for psychological explanation are uniformly helpful to reliabilist theories of justification" (Conee and Feldman 1998, p. 17; see also Baergen 1995).[5] However, in the preceding sections, we have delivered precisely such an account.

7 Concluding Remarks

We have discussed how agents' process tokens, instantiating particular types, do, in fact, to a large extent, converge, albeit that the processes involved are complex. We have further argued that a unified interpretation of the generality problem is plausible, where scientific types are seen to incorporate contextual and common sense features. In order to identify the process type in a given case, the best we can do may be to inquire into the contextual, evolutionary, developmental, social, and cultural factors relevant to the agent's background experience.

Our discussion of the generality problem has been principled, epistemically plausible, and true to the spirit of reliabilism. Token belief-forming processes instantiate a uniquely "right type" of the biological agent's cognitive architecture (including relevant embodied features) and background experience, consisting in processes of attractor-guided neural activation.

Acknowledgments We want to thank Martin L. Jönsson for numerous suggestions, and for him being exceedingly charitable with his knowledge and time. Thanks to James Beebe and our anonymous reviewers for comments.

Funding Open access funding provided by Lund University. The authors gratefully acknowledge support from Makarna Ingeniör Lars Henrik Fornanders fond, grant numbers FO2019-0005; FO2020-0004 and Stiftelsen Elisabeth Rausings minnesfond: forskning, grant number RFh2019-0226.

[5] Baergen (1995) argues that categorizations take place at lower levels, which then become more fine-grained on higher levels, and that this ability is what makes us able to identify relevant types.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Adler, J., & Levin, M. (2002). Is the generality problem too general? *Philosophy and Phenomenological Research, 65*(1), 87–97.

Alston, W. P. (1995). How to think about reliability. *Philosophical Topics, 23*(1), 1–29.

Alston, W. P. (2005). *Beyond "justification": Dimensions of epistemic evaluation*. Ithaca: Cornell University Press.

Baergen, R. (1995). *Contemporary epistemology*. Fort Worth: Harcourt Brace College Publishers.

Beebe, J. R. (2004). The generality problem, statistical relevance and the tri-level hypothesis. *Noûs, 38*(1), 177–195.

Buzsáki, G. (2006). *Rhythms of the brain*. Oxford: Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204.

Clark, A. (2015). *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Comesaña, J. (2006). A well-founded solution to the generality problem. *Philosophical Studies, 129*(1), 27–47.

Conee, E., & Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies, 89*(1), 1–29.

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive Science, 33*(5), 739–742.

Feldman, R., & Conee, E. (2002). Typing problems. *Philosophy and Phenomenological Research, 65*(1), 98–105.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: beyond "heuristics and biases.". *European Review of Social Psychology, 2*, 83–115.

Goldman, A. I. (1979). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge* (pp. 1–23). Dordrecht: Reidel.

Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge: Harvard University Press.

Goldman, A. I. (2016). Reply to Olsson. In B. P. McLaughlin, and H. Kornblith (eds.), *Goldman and his critics* (pp. 197-199). (Philosophers and their critics; Vol. 16). Chichester: Wiley-Blackwell.

Goldman, A. I. (2017). What can psychology do for epistemology?: revisiting epistemology and cognition. *Philosophical Topics, 45*(1), 17–32.

Goldman, A., and Beddor, B. (2016). Reliabilist epistemology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), URL = <https://plato.stanford.edu/archives/win2016/entries/reliabilism/>.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review, 109*(1), 75–90.

Heller, M. (1995). The simple solution to the problem of generality. *Noûs, 29*(4), 501–515.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Horst, S. (2016). *Cognitive pluralism*. Cambridge: MIT Press.

Jönsson, M. L. (2013). A reliabilism built on cognitive convergence: an empirically grounded solution to the generality problem. *Episteme, 10*(3), 241–268.

Jönsson, M. L. (2015). Linguistic convergence in verbs for belief-forming processes. *Philosophical Psychology, 28*(1), 114–138.

Kampa, S. (2018). A new statistical solution to the generality problem. *Episteme, 15*(2), 228–244.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (Eds.). (2013). *Principles of neural science* (5th ed.). New York: McGraw-Hill, Health Professions Division.

Kinzler, K. D., & Spelke, E. S. (2007). Core systems in human cognition. *Progress in Brain Research, 164*, 257–264.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to western thought*. New York: Basic Books.

Lee, C. J. (2007). The representation of judgment heuristics and the generality problem. *Proceedings of the Annual Meeting of the Cognitive Science Society, 29*(29), 1211–1216.

Markman, E. M., Horton, M. S., & McLanahan, A. G. (1980). Classes and collections: principles of organization in the learning of hierarchical relations. *Cognition, 8*(3), 227–241.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Mizumori, S. J. (2013). Context prediction analysis and episodic memory. *Frontiers in Behavioral Neuroscience, 7*, 132.

Mobus, G. E., & Kalton, M. C. (2015). *Principles of systems science*. New York: Springer.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge: A Bradford Book, MIT Press.

Olsson, E. J. (2016). A naturalistic approach to the generality problem. In B. P. McLaughlin, and H. Kornblith (eds.), *Goldman and his critics* (pp. 178-199). (Philosophers and their critics; Vol. 16). Chichester: Wiley-Blackwell.

Peirce, C. S. (1877). Illustrations of the logic of science: the fixation of belief. *Popular Science Monthly, 12*(November), 1–15.

Quine, W. V. O. (1969). Epistemology naturalized. In *Ontological relativity and other essays* (pp. 69–90). New York: Columbia University Press.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350.

Rysiew, P. (2017). Naturalism in epistemology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), forthcoming. URL = <https://plato.stanford.edu/archives/spr2017/entries/epistemology-naturalized/>.

Smith, D. M., & Bulkin, D. A. (2014). The form and function of hippocampal context representations. *Neuroscience & Biobehavioral Reviews, 40*, 52–61.

Sporns, O. (2011). *Networks of the brain*. Cambridge: The MIT Press.

Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge: The MIT Press.

Webb, T. W., & Graziano, M. S. (2015). The attention schema theory: a mechanistic account of subjective awareness. *Frontiers in Psychology, 6*, 500.

Wunderlich, M. E. (2003). Vector reliability: a new approach to epistemic justification. *Synthese, 136*(2), 237–262.

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

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

LUND
UNIVERSITY

Department of Philosophy
Faculty of Humanities
Lund University

ISBN 978-91-89874-02-2

9 789189 874022