5

Making Naturalised Epistemology (Slightly) Normative

MARCIN MIŁKOWSKI

1 Engineering Epistemological Normativity

The standard objection against naturalised epistemology is that it cannot account for normativity in epistemology (Putnam 1982; Kim 1988). There are different ways to deal with it. One of the obvious ways is to say that the objection misses the point: It is not a bug; it is a feature, as there is nothing interesting in normative principles in epistemology. Normative epistemology deals with norms but they are of no use in practice. They are far too general to be guiding principles of research, up to the point that they even seem vacuous (see Knowles 2003).

In this chapter, my strategy will be different and more in spirit of the founding father of naturalised epistemology, Quine, though not faithful to the letter. I focus on methodological prescriptions supplied by cognitive science in re-engineering cognitive architectures. Engineering norms based on mechanism design were not treated as seriously as they should in epistemology, and that is why I will develop a sketch of a framework for researching them, starting from analysing cognitive science as engineering in section 3, then showing functional normativity in section 4, to eventually present functional engineering models of cognitive mechanisms as normative in section 5. Yet before showing the kind of engineering normativity specific for these prescriptions, it is worthwhile to

Beyond Description: Naturalism and Normativity.
Marcin Miłkowski and Konrad Talmont-Kaminski (eds.).

review briefly the role of normative methodology and the levels of norm complexity in it, and show how it follows Quine's steps.

Quine insisted on reducing normativity to engineering as taught at technical universities, even ethical normativity (Quine 1994). He seemed to accept (Q1) or (Q2):

(Q1) Engineering helps to explain what normativity is.

(Q2) Engineering helps to reduce normative propositions to practical/technical knowledge.

(Q2) seems to be more coherent with Quine's position, as he is not in general interested in explaining previous theories or frameworks in terms of his reductive proposals. The idea is to reduce meaning to behaviour, and replace mentalism with behaviourism, and not to explain why mentalist talk is justified. The explanatory task hinted in (Q1) does not exclude (Q2) but seems less entrenched in his project.

It is still an open question how far one can go with substituting normativity with engineering principles. I doubt that ethical theories are just like engineering and should be developed at technical universities by engineers, as Quine bluntly claimed. For most moral philosophers today, naturalised ethics is supposed to save the phenomenon of normativity and explain it rather than to replace it with engineered principles and explain it away. A Nietzschean kind of moral philosophy that is committed to naturalism and sceptical about the validity of any received moral principles is rare, as naturalism is hardly committed to overarching scepticism. Partial scepticism about the validity of the received view is, however, one of the reasons why naturalism remains attractive for many philosophers. In epistemology, naturalists can be sceptical about norms in epistemology or justification as the most important theme in theory of knowledge. Yet Quine's suggestion is antisceptical in this regard: "Insofar as theoretical epistemology gets naturalised into a chapter of theoretical science, so normative epistemology gets naturalised into a chapter of engineering: the technology of anticipating sensory stimulation" (Quine 1992: 19).

Pace Quine, anticipating sensory stimulation is not the essential aspect of normative epistemology, as cognition is not only about sensory stimulation and external behaviour, as behaviourists thought. Moreover, the goal of science should not be conflated with prediction (described as anticipation of sensory stimuli) while science is also interested in explanation, understanding and description. Quine could be right about engineering but knowledge engineering is not only about anticipating sensory stimuli; it is about generating knowledge, but not necessarily predictive knowledge. So the claim inspired by Quine becomes (Q*):

(Q*) By using engineering knowledge, it is possible to define normative criteria for naturalised epistemology.

The term "normative criteria" is deliberately kept abstract enough to cover various kinds of norms. These are not only the principles of cognitive action that are endorsed by agents but also the guidelines for developing cognitive agents, either by natural selection, or artificially. Another kind of normative criteria is connected to those conditions of interaction among groups of cognitive agents and with environment that prove to be fruitful for development of knowledge. So there are at least three categories of norms: (1) endorsed by individuals, (2) embodied by individuals and (3) embodied by interaction of individuals with each other and with their environment. They are all methodological.

Methodology does not help us develop norms that are valid categorically. As instrumental, they are valid conditionally. In other words, methodology does not supply us with the norms that are justified the way the moral obligations were supposed to be justified with the help of Kant's categorical imperative. It is rather on a par with the hypothetical imperative that supplies conditional norms such as: If you want to achieve X, you should do Y. As Quine would say, if you want to anticipate sensory stimulation X, you should do Y (see also Hookway 2002: 39). However, the antecedent of such conditionals does not have to be universally valid or true. In this regard, I agree with Knowles 2003 that obligatory universal normative principles are rare in epistemology, if they exist at all. This does not mean that there are no prescriptions in epistemology. Sometimes they are even more restricted than simple conditionals: A methodological prescription can suggest a method that fails in some cases, for example a heuristic method, or a method that is sometimes advisable but not strictly required.

2 Methodological Naturalised Epistemology

In naturalised epistemology, methodology seems to have been largely ignored. For example, neither reliabilism nor evidentialism seems to have any bearing for methodology. They just piggy-back on methodology but give nothing in return, being motivated solely by philosophical interest in reducing epistemic concepts to nonepistemic or by an attempt to define knowledge in a Gettier-proof way. This is not, however, based on any principled argument against normative methodology in sciences. In non-armchair epistemology, methodological questions are to be settled by scientists rather than with conceptual analysis alone but nothing prevents from integrating those questions into the research program of naturalised epistemology.

I suggest that nonfoundational epistemology can be methodological and as such normative. A nonmethodological normative epistemology (foundational or not) does not seem to make much sense for non-armchair naturalism, or otherwise it could not be of any use for science. Methodology cares about proper methods of acquiring and justifying knowledge by cognitive agents.

Naturalised epistemology can be analysed as operating at least on three levels of complexity:

- *cognitive system interaction* (fashions, traditions, groups, niches…)
- *personal* (beliefs, webs of beliefs, theories, desires…)
- *subpersonal* (perceptual subsystems, faculties, cognitive modules…)

The distinction between the personal and subpersonal level of description (Dennett 1969) is sometimes difficult to spell out precisely, yet I need it only for showing that naturalised epistemology can operate on many levels of complexity (the notion of level used in this chapter is the same as in Craver 2001). The additional cognitive system interaction level hints at developments connected with notions of embodiment and embeddedness in cognitive science.

Cognitive system interaction is a dimension of epistemology that is explored by science studies, sociology of knowledge, evolutionary epistemology and other disciplines that are not committed to methodological solipsism and/or methodological individualism. In my terminology, it covers also interpersonal processes. Interaction between individuals generates a new level of complexity that is different not only quantitatively but also qualitatively from individual cognitive achievements[1].

---

[1] Traditionally, sociology of knowledge has been treated as a forerunner of relativism or even irrationalism by philosophers of science, but it is just an empirical assumption that individual researchers always act like strictly rational and good-willing angels that care only about truth and the growth of knowledge, and not about their own careers and eco-

On the personal level, one can describe the capabilities of a cognitive agent to acquire knowledge, the ways she justifies her beliefs, etc. These descriptions require using intentional predicates, especially in the realm of individual webs of beliefs. The personal level is preferred by antipsychologist, nonnaturalist epistemology, as it is possible to analyse beliefs in a pure logical fashion, for example to criticise the logical structure of reasoning. Yet, logical hermeneutics deals with just one kind of idealisations of personal-level beliefs, and not always the most important kind: psychotic beliefs can be logically coherent but completely out of sync with reality.

On the subpersonal level, epistemology specifies the mechanisms within the cognitive agent that contribute to her personal-level cognitive abilities. In traditional epistemology, subpersonal properties were construed of as psychological (in terms of a philosophical, rational psychology). Nowadays, cognitive science deals with this level of explanation. As cognitive science is a multidisciplinary research program rather than a unified science, the predicates that describe subpersonal properties do not have to be psychological only but to span multiple levels and disciplines (for multiple-level explanations in neuroscience, see Craver 2007).

Antinaturalist and nonnaturalist philosophy denies the importance of the subpersonal level in epistemology, which is sometimes motivated by antipsychologism. At the same time, all epistemological theories, including antinaturalist epistemology, use the intentional idiom to speak about cognitive capabilities. So what antipsychologism denies is not the usage of intentional predicates such as "believe", "think" or "perceive" but citing nonintentional psychological facts in justifying knowledge. In epistemology, rationality, logical structures and justification count, and these concepts can be analysed without knowing any underlying psychological mechanisms in cognitive agents.

Interestingly, many modern philosophers, including especially Immanuel Kant, were analysing the underlying mental structures, by assuming that they are inaccessible to introspection, in order to vindicate the concepts of rationality and justification. Kantian categories and other theoretical mechanisms or entities such as transcendental apperception are on the subpersonal level. The way the category of causality operates is not the way the cognitive agent acts. According to antipsychologism,

nomic interests. At least in some cases, this assumption is evidently false. Not to mention that science is only possible not as an individual but a collective achievement.

this must be viewed as a commitment to psychologism, and implies a serious methodological error, or Kant's theory is to be reconstructed in terms of logical structures. I suggest otherwise.

In Kantian epistemology, the hypothesised subpersonal structures are *a priori*: Their use is justified independently of experience[2]. In cognitive science, empirical hypotheses about analogous structures and processes are being developed, and naturalised epistemology cheerfully endorses those hypotheses (if confirmed by empirical data).

The subpersonal level is necessary for any cognitive system to be cognitive. While it is possible to abstract from the way the cognitive personal level capabilities are realised on the lower level, and describe the logical structure of beliefs or theories, those capabilities are instantiated always and only if underlying lower level properties are also instantiated; higher level properties depend on the lower level properties of the system. Note that because there could be external factors on higher levels, such as other cognitive systems, time pressure, or tiredness, that may affect cognitive performance, higher level properties in most cases will never be fully reducible to lower level properties[3].

At the subpersonal level, the cognitive architecture that generates knowledge in a system is the most important factor. Note that on any analysis of the notion of knowledge (traditional, reliabilist, disjunctivist…), you cannot ascribe true knowledge to a dysfunctional cognitive system. In next sections, I will focus on the subpersonal level and its functional characteristics to show one of the kinds of normativity in naturalised epistemology.

The body of knowledge in scientific methodologies concerns mostly personal level achievements of cognitive agents, and as far as it is empirically valid, naturalised epistemology cannot question it. Environmental and other interaction-dependent factors, which are now being taken into account (if sometimes exaggerated) in embodied and embedded cognition theories, also play a role in cognitive processes. The norms on the personal level are most likely to be methodological prescriptions

---

[2] If you understand Kant's project nonfoundationally, you can see that a priori does not mean conceptual; for a nonfoundational reading of Kantian project, see Kitcher 1995 and Miłkowski 2007.

[3] This is because systems on all three abovementioned levels of organization are never fully isolated or completely autonomous, and they can be influenced from the outside.

to be followed consciously, and their normative character is just of the kind displayed by the hypothetical imperative.

In order to be a successful cognitive agent, you should behave the way other successful cognitive agents behave according to scientific methodology prescriptions. For example, a nonpractical norm of scientific methodology would be, as Knowles 2003 shows, "avoid contradiction", as you cannot effectively rule out contradiction in most real-world cases due to the immense computational complexity of the task (in decidable logical systems; it is not feasible in incomplete systems). However, more detailed norms, like "change beliefs if you find a deep contradiction" or "treat contradiction as a nondirect proof" are more viable in some cognitive niches, like mathematics. Another viable norm at the personal level for a modern scientist would be "submit your papers to peer-reviewed journals".

On the interaction level, the norms are also instrumental. There are functional systems that interact and their interactions may become dysfunctional, so there could be functional normativity. Cognitive agents can communicate and share hypothetical imperatives. It is an open question whether conscious sharing and debating norms among cognitive agents should be accounted for as generating simply personal-level norms, or as involving larger systems (i.e., organizations such as research agencies) with their level of hypothetical imperatives. I leave the personal and interaction level questions mostly aside as they are outside of the scope of the current paper. However, norms on all levels stay instrumental and are not normative *sui generis*: they are reducible to hypothetical imperative conditionals that do not express obligations.

3 Cognitive Science and Evolutionary Theory as Engineering

As Dennett (1995) argues, cognitive science and evolutionary biology are engineering sciences, construed of along the lines of the "sciences of the artificial" in Herbert Simon's sense (Simon 1996). These sciences use the "design stance" to describe and predict phenomena. The difference between the standard engineering practice and these sciences is that the latter are focused mainly on reverse engineering, while nonreverse engineering is used to confirm the hypotheses about the reliability of the design being reverse-engineered. At the same time, there is a growing research community on artificial life and artificial intelligence, which are clearly closely linked with engineering, evolutionary theory and cognitive science.

I suggest that naturalised epistemology uses exactly the same stance as cognitive science to describe cognitive processes on the subpersonal level. The notion of "function" is not a physical notion, and it is not ascribed from the physical stance but from the design stance (Dennett 1971). Subpersonal structures are functional: the human mind is conceived of as a complex system of interacting faculties (or modules) that contribute to personal-level cognitive capabilities (for a vindication of massive modularity, see Carruthers 2006).

The task of cognitive science is to discover the functional structures and processes of mind on the subpersonal level, and these structures are of interest for normative naturalised epistemology. Most (if not all) naturalised theories of cognition should be translatable into engineering projects: if you know how cognition proceeds, you should be able to replicate the process in an artefact or explain why it cannot be replicated.

In other words, in order to build successful cognitive agents, you need to engineer and/or reverse-engineer their cognitive architecture. By investigating the individual components of the architecture and their interaction, you can see which solutions are functional and which are not. If you know why a system fails to function normally, and know the way to fix it, you can repair it. After finding a viable way to extend its cognitive abilities, you can supply the system with new cognitive components that will enhance its cognitive performance. All three cases: discovery of the architecture, maintaining the performance of the architecture, and extending the architecture, imply a specific kind of normativity. This normativity is embedded in the engineering of the cognitive architecture.

4 Cognitive Architecture Functions in Normative Epistemology

Subpersonal level cognitive architecture is described from the design stance. Design stance specifications provide descriptions of functions or functional complexes. The notion of function has been discussed in philosophy of science for several decades, and at least three kinds of function concepts have been proposed: causal role based, etiological, and structural. The causal role account ascribes function to a system component that contributes causally to the system level capability (Cummins

1975). The etiological account ascribes function to a system component that has been selected in the system because it contributed causally to the system capability in the past (Wright 1973; Millikan 1984). The structural account ascribes function to a system component when it serves a function in the system design (Krohs 2004, Krohs 2009) or contributes to the system autonomy (Bickhard 2000, Bickhard 2004). The design-based notion requires that a component of the system be ascribed function not just because it has a causal role or was inherited but because it is selected by the design for this function. The selection process can be natural or artificial.

It is often noted that functional attributions are normative in a very specific sense: something may fail to function properly, and this failure is used for evaluative purposes. The causal account of function fails to make the distinction between function and dysfunction: if a TV set breaks, the causal chains still occur and they contribute to the system capability of keeping the screen black, for example. The etiological account fails for first tokens of the system, as they have no selection history that would specify the components. For this reason, the prototypes of AI systems will always be dysfunctional. Autonomy account would also make partial AI systems dysfunctional: as long as you do not have the model of the whole iguana, no part of the iguana can be functional. But because the research starts with relatively isolated subsystems that are not fully autonomous, it would be useful to be able to distinguish the properly functioning systems from the broken ones. This is what the design-based account enables (Krohs 2009). It is a happy coincidence that the design account uses the notion of design also used in the notion of the design stance. In what follows, I will assume that design stance functions be analysed in a design-based account.

A design-stance description refers to a functional structure (design) of the system. Cognitive architecture description provides criteria for evaluation of the functioning of a cognitive system and its subsystems. If something essential in the architecture is missing, system becomes dysfunctional, and fails to fulfil cognitive functions. The notion of architecture, i.e., the abstract specification of the system component interactions, is equivalent to the notion of the system design.

Functional normativity is instrumental. If a TV set stops functioning, it is not a failure to fulfil a moral prescription. It is just useless as a TV set. This kind of normativity can be analysed either in terms of means and ends (the end is to watch television on the device, the means is the architecture that works in home environment), or in terms of instrumental values. In both cases, it is a derived kind of normativity, subsumable under the hypothetical imperative. But we do not need any more normativity on the subpersonal level in naturalised epistemology.

5 Artificial Cognitive Systems As Models for Epistemology

In research on cognition, one can take a biologically inspired approach (reverse engineering) or artefact-inspired approach (engineering). These approaches can intermingle, as artefacts are often also biologically inspired, like wings of a plane. But there are artefacts that have no counterparts in the evolutionary world, such as wheels, that cannot be evolved for morphological reasons but remain–in their niche–a very good technical solution. It remains an open question whether artificial systems that are able, for example, to prove theorems in a way that is too difficult for a human being (like the four colours theorem), are minimally cognitive. We may expect, however, that future cognitive systems will have capacities that are more like wheels than like wings.

These wheel-like cognitive capacities are at least as important in naturalised epistemology as research on biological cognition. Building cognitive systems that are not modelled on animal or human cognition has always been the goal of epistemological theories. Specifically, one can set out to sketch the overall design specifications for a cognitive robot–these would serve roughly the same role as the traditional notion of the transcendental subject. For that purpose, conceptual arguments will not suffice. Naturalised epistemology must consult engineering sciences such as AI and cognitive science.

So joining the forces between philosophical analysis of cognition per se and engineering of cognitive systems can lead to android epistemology, as Clark Glymour dubbed it–to epistemology that can analyse nonbiological cognitive systems (see Ford, Glymour & Hayes 1995; Ford, Glymour & Hayes 2006). They could be arbitrarily stronger or weaker than natural cognitive systems. Just by juxtaposing those artificial models with natural agents, one could try to see the really cognitive aspect of their actions.

The research on artificial cognitive architectures has several benefits. First of all, it is ethical to test which dysfunctions are caused by switching off some modules in a simple artificial cognitive system. This is impossible with research on human beings, and it can be argued that such experiments on animals are also too cruel. A second benefit is that translation of abstract models into engineering specifications requires that the conceptual vagueness be replaced with specific solutions. Thirdly, unexpected consequences of idealised models are discovered: the frame problem can be thought of as an unexpected implication of the propositional model of cognition that reduces cognition to reasoning in classical logic. Even approaches that were mildly hostile to engineering in investigating cognition, like Heidegger's, can be turned into requirements for making cognitive systems as embedded in their environments. In the process, the hermeneutical model of interpreting Heidegger's *Sein und Zeit* is replaced with actually improving upon his ideas (see for example Wheeler 2005).

Another benefit is that it becomes testable whether a given model of cognition works or not. Though initially it can be unclear why a given system is dysfunctional, after comparing different implementations of the similar architecture, one can see if it can really work. This is of course impossible for armchair conceptual analysis.

The cognitive science on the subpersonal level can supply normative principles for cognitive architectures, and these normative principles, which would generally determine what is minimally required for certain cognitive capacities, are available for use in naturalised epistemology. Similarly, personal-level intentional requirements of rationality as studied by methodology and psychology supply normative principles. They are instrumental as well–if they do not involve any moral obligation. The interaction level is the same in this regard, as it uses both kinds of principles. To wit, naturalised epistemology can be normative if it uses detailed results of cognitive science, science methodology, psychology and other sciences. The normative principles in epistemology are not however vague and general norms like "avoid contradiction", but specific and constrained heuristic principles of design.

There is no general and effective algorithm for discovering the laws of nature or deducting all possible true propositions, as Tarski and Gödel proved. Hume's problem of induction will remain unsolved for limited beings. So there is no hope that a single methodological prescription would say what to do to get all true knowledge as it would presuppose that there is an effective algorithm to get them. At the same time, there are effective algorithms that deal only with discovering some laws in a limited domain or deducting some true propositions (if the system is incomplete). So there might be constrained algorithms with a limited scope of application in normative epistemology, but there is no hope for a groundbreaking normative philosopher's stone.

References

Bickhard, Mark H. (2000). "Autonomy, Function, and Representation". *Communication and Cognition - Artificial Intelligence* 17(3–4), pp. 111–131.

Bickhard, Mark H. (2004). "Process and Emergence: Normative Function and Representation", *Axiomathes - An International Journal in Ontology and Cognitive Systems* 14, pp. 135–169.

Carruthers, Peter (2006). *The Architecture of the Mind. Massive Modularity and the Flexibility of Thought*. Oxford: Clarendon Press.

Craver, Carl F. (2001). "Role Functions, Mechanisms, and Hierarchy". *Philosophy of Science* 68, pp. 53–74.

Craver, Carl F. (2007). *Explaining the Brain*, Oxford: Oxford University Press.

Cummins, Robert (1975). "Functional Analysis". *Journal of Philosophy*, 72 (20), pp. 741–765.

Dennett, Daniel C. (1971). "Intentional systems". *Journal of Philosophy* 68 (February), pp. 87–106.

Dennett, Daniel C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* New York: Simon & Schuster.

Ford, Kenneth M., Glymour, Clark, Hayes, Patrick, eds. (1995). *Android Epistemology*, Cambridge, Mass.: MIT Press.

Ford, Kenneth M., Glymour, Clark, Hayes, Patrick, eds. (2006). *Thinking about Android Epistemology*. Cambridge, Mass.: MIT Press.

Hookway, Christopher (2002). "Naturalism and Rationality". In *Quine. Naturalised Epistemology, Perceptual Knowledge and Ontology*, ed. by Lieven Decock and Leon Horsten, Poznań Studies in the Philosophy of the Sciences and the Humanities. Amsterdam: Rodopi.

Kim, Jaegwon (1988). "What is 'Naturalised Epistemology'?". *Philosophical Perspectives* 2, pp. 381–405.

Kitcher, Patricia (1995). "Revisiting Kant's Epistemology: Skepticism, Apriority, and Psychologism". *Nous* 25 (3), pp. 285–315.

Knowles, Jonathan (2003). *Norms, Naturalism and Epistemology: The Case for Science without Norms*. Palgrave Macmillan.

Krohs, Ulrich (2004). *Eine Theorie biologischer Theorien*, Berlin: Springer Verlag.

Krohs, Ulrich (2009). "Functions as Based on a Concept of General Design", *Synthese* 166 (1), pp. 69–89.

Millikan, Ruth G. (1984). *Language, Thought, and Other Biological Categories. New Foundations for Realism*. Cambridge, Mass.: MIT Press.

Miłkowski, Marcin (2007). "Siemek, dialektyka, rzeczywistość. O transcendentalnej filozofii społecznej". *Przegląd Filozoficzno-Literacki* 16 (1), pp. 99–113.

Putnam, Hilary (1982). “Why Reason Can’t Be Naturalised”. *Synthese* 52 (1), pp. 3–23.

Quine, Willard Van Orman (1992). *The Pursuit of Truth*. Revised edition. Cambridge, MA: Harvard University.

Quine, Willard Van Orman (1994). *In Conversation – W. V. Quine. The Dennett panel* (VHS cassette). London: Philosophy international.

Simon, Herbert (1996). *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press.

Wheeler, Michael (2005). *Reconstructing the Cognitive World. The Next Step*. Cambridge, Mass.: MIT Press.

Wright, Larry (1973). “Functions”. *The Philosophical Review* 82 (2), pp. 139–168.