Natural Sources of Normativity

Wayne Christensen

Address: Konrad Lorenz Institute for Evolution and Cognition Research; Adolf Lorenz Gasse, 2; A-3422, Altenberg, Austria
E-mail: wayne.christensen@gmail.com
Telephone: (+43) 2242-32390-13
Fax: (+43) 2242 323904

Abstract

Normativity is widely regarded as being naturalistically problematic. Teleosemantic theories aimed to provide a naturalistic grounding for the normativity of mental representation in biological proper function, but have been subject to a variety of criticisms and would in any case provide only a thin naturalist platform for grounding normativity more generally. Here I present an account that identifies a basic form of valuational normativity in autonomous systems, and show how the account can be extended to encompass key aspects of the normativity of functions and practical reasons.

Keywords

Normativity, naturalism, autonomous systems, functions, reasons, persons
1. Introduction: normativity and naturalism

Normativity is paradigmatically a matter of right and wrong, good and bad. Philosophical work on normativity seeks to understand the nature of normative claims, the nature of justification for such claims, and the fundamental sources of normativity. One common view is that there is nothing in the natural world, accessible by scientific means, which grounds normative claims. The most influential arguments to this effect are due to Hume and G. E. Moore: Hume argued that no normative conclusion can be validly derived from descriptive premises (Hume, 1978), whilst Moore’s ‘open question’ argument asserts that any attempt to identify a normative property (e.g., goodness) with a natural property (e.g., pleasure) is always open to doubt, thus showing that conceptually the two cannot be identical (Moore, 1971). The popularity of this view is probably due to a more complex set of influences than just the force of these arguments, however. Lurking in the background are a pair of ideas that tend to work hand-in-hand: on the one hand, the idea that modern science replaced Aristotelian teleology with mechanistic explanation, and on the other, the idea that normativity is a very special feature of human agency, linked to consciousness and perhaps the capacity for reflection.

Whatever the exact reasons, it is often thought that naturalistic theory should not stray over the putative fact/value boundary. Yet naturalist theory in this mode must overcome a major obstacle, which is that normativity seems to be an endemic and very important feature of human agency. Not only moral agency, but cognitive agency more broadly. Representations can misrepresent, words can be used wrongly, people can leap to irrational conclusions, and they can act unwisely. If adopting a scientifically based perspective means giving up normativity, this is giving up a lot. Naturalists practicing an austere norm eliminativism aim to show that these phenomena can be understood without appeal to normative concepts, despite appearances to the contrary, but it is not unreasonable to doubt that the project can succeed. Normative eliminativism may be an unnecessary straightjacket, however. Here I will sketch a naturalist approach that follows Aristotle
in recognizing relatively rich forms of normativity in living systems. Specifically, it sees normativity as inherent in the organization or form of living systems, specifically in the form that generates their unity and hence explains their existence.

The most immediate point of comparison for this account is the etiological theory of normative function. The approach to functional normativity advocated here differs in fundamental ways with the etiological theory, and indeed with most other contemporary accounts of normative function, inasmuch as it begins with a different explanatory agenda. On the usual conception the task of function theories is to explain how functions are assigned to parts, whereas the approach taken here instead focuses on explaining value in relation to systems, and much of the emphasis is on identifying the relevant class of systems. This is done by means of a theory of the fundamental organization of living systems. The basic idea is not especially novel: as noted, it treads in the footsteps of Aristotle, and there are a variety of contemporary theories that attempt to give an account of the organization of living systems, which often assume that functional normativity pertains to these systems in virtue of their organizational structure.1 Here I attempt to flesh out the intuition in a way that relates it to a broader understanding of normativity.

2. Normativity: some basic distinctions

Before proceeding further it will help to sketch out the nature of normativity in a little more detail. This cannot be done in an uncontroversial way, but the following distinctions capture at least approximately some of the major forms of normativity that have been discussed (see e.g. Darwall, 2001; Schroeder, 2008; Glüer and Wikforss, 2009). The initial description given above associates normativity with evaluation and prescription, but some have identified a kind of normativity referred to as ‘descriptive.’ Descriptive or ‘non-evaluative’ norms are such that it is possible to

---

1 These theories usually focusing on self-maintenance and/or self-production; see e.g. Schrödinger (1944), Maturana and Varela (1980), Bickhard (1993), Christensen and Hooker (2000), Christensen and Bickhard (2002), Kauffman (2003), Moreno et al. (2008), Barandiaran et al. (2009), Toepfer (2011).
specify conformance or departure from the norm, but there is no reason from this alone to think that there ought to be conformance to the norm, or that nonconformance is bad. Etiological proper functions (discussed in the next section) are thought by most proponents to have descriptive normativity (Neander, 2009). If we include such non-evaluative norms within the realm of the normative then the minimal kind of normativity may simply involve some kind of non-arbitrary framework allowing comparison between actual and alternative states. There is room to doubt that this is sufficient for normativity, but it may at least be necessary.

In the case of ‘evaluative’ normativity the comparison between actual and alternative states takes the form of a valenced assessment. ‘Valuation’ (traditionally addressed by axiology) involves assessments such as ‘good,’ ‘better than,’ and ‘worse than.’ ‘Prescription’ (traditionally addressed by deontic theory) specifies what ought or ought not to happen, with the biblical commandment ‘thou shalt not kill’ being a paradigm example of a (candidate) prescriptive norm. ‘Constitutive norms’ specify rules which must hold if something is to exist, such as the rules of a game like chess. They are per se non-evaluative, though they can inform evaluations in conjunction with other information, such as an agreement (perhaps tacit) to play by the rules. With regard to games and other activities we can further distinguish ‘performance norms,’ concerned with how well the game or activity is conducted, with winning, losing and ‘playing well’ counting as paradigmatic performance norms.

3. Etiological theories of normative function

Since the mind is often thought to be entirely or at least substantially functional in nature, theories of normative function are an obvious starting point for developing naturalist accounts of the normativity of cognitive phenomena. The teleosemantic program takes this route, attempting to ground the normativity of mental representation in biological function (Millikan, 1984; Papineau, 1984). Causal theories of representation, such as that of Dretske (1981), attempt to explain to
explain mental representation in terms of causally based correlations. Thus, activity in a toad’s retina is correlated with events in the world, and thereby represents those events. The familiar problem is that understanding representation in terms of causal correlation leaves no room for misrepresentation, because correlations either exist or they do not, they cannot be ‘false.’ But there do seem to be false representations. Teleosemantics offered a solution by appealing to an etiological theory of normative function. It specifies what a representation is supposed to represent in terms of the ‘proper function’ of the mechanism doing the representing; thus, toads will respond to a long dark horizontally moving stimulus as if it is a worm, and it seems reasonable to think that this is what the detection system in their brain is supposed to indicate. In the lab they respond to artificial stimuli created by the scientist, but in these cases they are misrepresenting. The etiological theory of normative function explains proper function in terms of natural selection: the proper function of an item is the function it is adapted to perform. This has been an appealing pathway for a naturalist account of normativity because normativity is explained by appeal to a natural phenomenon (evolutionary adaptation) that is relatively well understood, clearly of great importance, and is intuitively normative (as the putative basis of ‘biological design’).

Nevertheless, for some progress with this approach has not met expectations (e.g. Godfrey-Smith, 2006). Specific difficulties in the analysis of representational content need not concern us here, however some of the deeper and thornier issues stem from the basic source of normativity. As noted above, the normativity of etiological functions is supposed to be descriptive rather than evaluative. Thus, on Millikan’s account the proper function of a heart is to do what ancestor hearts did that made them the target of selection. But identifying this putative proper function will not allow us to conclude that this heart now ought to do what it’s ancestor hearts did, or that it is bad if it does not. By avoiding evaluation the theory evades Hume and Moore, however the result is a very thin and somewhat peculiar kind of normativity. Deviance from an ancestral state subject to positive selection is called ‘malfunction,’ but malfunction defined this way is not really ‘mal’: there is
nothing inherently bad about it (cf. Ferguson, 2007). Indeed, an etiologically defined malfunction
may be functionally advantageous in the current context. It would be clearer and more accurate to
replace the terms ‘proper function’ and ‘malfunction’ defined according to etiological theory with
technical labels that have no evaluative associations. For instance, we could replace ‘proper
function’ with ‘AS-function’ (for ‘ancestrally-selected function’), and replace ‘malfunction’ with
‘C-function’ (for ‘changed function’). With these substitutions the etiological theory no longer
appears normative, which suggests that it is getting illegitimate normative ‘oomph’ by means of
evocative labels. Without this oomph the grip provided by the theory is unconvincing: as we saw,
the etiological theory is the grounding point for the teleosemantic account of misrepresentation, but
misrepresentation defined this way is not really malfunctioning, or ‘incorrect’ functioning, it is just
different functioning. If we think there is something genuinely incorrect about misrepresentation
then we need more resources than the etiological account is providing.

The etiological approach is pseudo-prescriptivist in the sense that it gives something of the flavor of
prescriptivity without the actual prescriptivity. It does not aim to explain valuation, nor does it
support valuational assessment for the reasons just given. Yet, at least on first appearances,
biological functioning seems to involve valuational normativity: an organ can function well or
poorly, and an organism can be healthy or sick. Since etiological theory has nothing to say about
these kinds of phenomena (again, talk of ‘malfunction’ is deceptive), it would seem to be at best
incomplete insofar as it is supposed to account for functional normativity in biological systems.
Certain theoretical alternatives have been rejected on grounds that will be considered next, but for
reasons that are unclear the literature on normative function has been fixated on a prescriptivist
model of functional normativity, according to which functional normativity consists in the
possession by an item of a proper function, which the item is in some sense ‘supposed’ to perform
(Wright, 1973). The issues of normative perspective and functional value have been largely neglected.

4. Autonomous systems and normative function

Proponents of the etiological account of normative function have been drawn to it in part because they are skeptical of system theories of function. Cummins (1975) is thought to have provided the canonical account of system-based analysis of function, and on his account functional analysis is interest relative, in the sense that more or less anything can be given a systems functional analysis and ascribed function on that basis. Just as we can analyze the propensity of hearts to pump blood, we can analyze biologically irrelevant relations such as the contribution of the heart to body mass (Sober, 1993), or the propensity of mice to explode in space (Millikan, 1989). A particularly crucial claim is that the boundaries of a biological system cannot be identified on the basis of causal relations; only by identifying proper functions can a biological system be individuated in a principled way (Millikan, 1999). The importance of this claim is that it identifies a putative in-principle limitation. Cummins did not attempt to give a principled account of system individuation or normative function analysis, but this is not a reason to think it cannot be done. However if biological systems cannot be causally individuated that is a reason to think that any purely systemic analysis must be arbitrary in focus.

There are compelling reasons to reject the claim, however. As a matter of epistemology it cannot be the case that the principled identification of biological systems depends on the prior identification of etiological proper functions, because a theory of the adaptive origins of a trait is generated by

---

2 The problem of pseudo-normativity is not specific to the etiological theory; it arguably afflicts any account that tries to assign proper functions without a well-founded normative perspective. Thus, the system-based account of proper functions proposed by Schlosser (1998), and the development-based account of design (and proper function) by Krohs (2009), are both vulnerable to the type of criticism here leveled at the etiological account: the putative proper functions are not actually normative, and the use of the evaluative terms ‘proper function’ and ‘malfunction’ is misleading. Schlosser (1998) offers a system-based theory of normative function, but his account is structurally similar to the etiological account in that it aims to explain proper functions, and it doesn’t give an account of normative perspective or functional value. Krohs (2009) treats development rather than evolutionary history as the source of design for biological systems, and he considers this design to be non-intentional. One worry is whether the concept of non-intentional design really makes sense, but even if we accept the idea it will not support any substantive normativity.
analyzing the effect of the trait on the organism’s ability to survive and reproduce in the ancestral environment in which it appeared (Griffiths, 1993; Stotz and Griffiths, 2001). As a matter of ontology we should expect that living systems are physically individuated, because if they were not then there would be no physically distinct entities for natural selection to select over. If there are physical principles of individuation they can be used to supplement the systems framework for functional analysis that Cummins describes, providing a non-arbitrary basis for functional analysis.

This is the approach taken by Christensen and Bickhard (2002), who argue that an account of autonomy can serve as the basis for a suitable theory of individuation and system organization.³ This account has an Aristotelean flavor because it relates the organization systems to their unity and existence conditions. In a broad sense the organization of a system is just how it is arranged, and a basic question to ask for natural systems is how their organization is related to their existence. Often, many aspects of a system’s organization make little difference to its ongoing existence. Consider a collection of rocks scattered across the floor of a dry cave: the rocks have many different shapes and sizes, but for the most part the differences in shape, size and location will not affect the ongoing existence of the rocks in this very stable environment. On the other hand, on an open plain differences in shape, size and location will tend to have a stronger effect on rock existence because of their effects on weathering processes.

For some systems there is a very special relationship between their organization and their existence because (unlike rocks) they actively construct the conditions which give them unity and ongoing persistence. The concept of autonomy is intended to capture this idea, and according to the analysis of autonomy given by Christensen and Bickhard a system is autonomous if it tends to generate the conditions for its persistence, and if it has infrastructure that contributes to this self-maintenance.

³ The particular account of autonomy they use is previously developed in Christensen and Hooker (2000a, 2000b). Bickhard’s account of self-maintent systems is described in Bickhard (1993, 2000). More generally, this theory of autonomy is one of a family of theories that aim to give a systemic account of living systems – see footnote 1 above. Autonomy as it is used in this context must of course be distinguished from personal autonomy (Buss, 2008).
Infrastructure here refers to persistent, relatively stable structure that shapes more dynamic system-maintaining processes, with the cell membrane of living cells being a paradigm example. The infrastructure requirement rules out simple positive feedback systems, and indeed more complex feedback systems like tornados, which are self-maintaining but lack infrastructure that supports self-maintenance.

Establishing a principled basis for system individuation is a crucial anchor point for an account of normative function, because normative evaluations of function can be made relative to system identities. That is, functional relations can be assessed in terms of their effect on the system. The contribution of the heart to body mass clearly has little significance for the organism, whereas the contribution of the heart to fluid transport has profound importance. Christensen and Bickhard (2002) develop an account of normative function of this kind. It is not intended to explain how parts ‘have’ functions that they ‘should’ perform, it is instead intended to provide a valuational account of functional relations relative to the system as a whole.

Normative properties such as benefit and dysfunction are characterized. An item is beneficial for an autonomous system if it contributes positively to the autonomy of the system, and this can be so regardless of the whether the item ‘has’ the function of making this contribution. Similarly, if a system is autonomous it will be composed of a network of interdependent processes, and we can understand dysfunction in terms of these interdependencies. If the heart stops beating then there will be a cascade of failures as physiological processes that depend on fluid transport cease to function, leading to the death of the organism. The dysfunction here is systemic – a property of the pattern of network dependencies – and as such not attributable to the heart in isolation. If an alternative mechanism for fluid transport appears, such as an artificial heart, the dysfunction goes away. Again, these network dependencies can be analyzed quite independently of what functions ‘belong’ to the various parts.
One line of response the etiologist might take is that these kinds of network analyses are merely descriptive. Mirroring the critique of the etiological account given above, it could be said that terms like ‘benefit’ and ‘dysfunction’ are a colorful and misleading way of describing purely physical relations. We can say that a given item is or is not affecting the system in some way, but without the etiological account there is no basis for comparing actual performance against a normative benchmark. The etiological account provides at least this basic form of normativity by means of the distinction between what an item does and what it is for, and it is unique amongst naturalistic forms of function ascription in doing so.

This line of criticism is not strong, however. The autonomous systems account does provide a basis for normative comparison: using the autonomy of the system as a whole as the grounding point, we can compare actual state with alternative states that would be better or worse. Moreover, in this kind of analysis words like ‘benefit’ and ‘dysfunction’ are not misleading. One way to gauge this is by the fact that it is not possible to replace them with technical terms that have no evaluative content. We can certainly replace the words, but if we leave out the evaluative content we lose information. Thus, the systems account has a firmer normative basis than the etiological theory: it uses evaluative concepts in an informative way.

Another possible line of criticism is that autonomous systems do not have a legitimate normative perspective. We can talk about certain things being ‘good’ or ‘bad’ for these kinds of autonomous systems, but this is ultimately just as empty as talking about certain things as being ‘good’ or ‘bad’ for a rock. We can imaginatively think of the breaking of a rock as bad for the rock, but rocks do not really have the kind of normative perspective that warrants such evaluation. An initial response to this criticism is to point out that according to the theory autonomous systems are causally special in a way that makes them unlike rocks. Autonomous systems are organized such that they tend to be
self-perpetuating, and they have infrastructure which supports this self-perpetuation. We cannot properly understand the causal structure of these systems if we do not recognize that they are organized in a way that achieves self-perpetuation. And to understand this self-perpetuation at a more fine-grained level we must characterize the relations between the persistence of the system as a whole, and the constituent structures and processes.\(^4\) Thus, the use of evaluative concepts is not simply an imaginative projection, it is required to properly characterize the causal structure of these kinds of systems. Moreover, in many cases the infrastructure possessed by these systems is regulative: it repairs, avoids, seeks, etc. In a limited but significant way living systems are doing their own evaluation, which is a persuasive reason for treating them as having a genuinely normative perspective. The etiologist may point out that these systems have infrastructure for self-perpetuation largely as a result of an evolutionary history. The autonomous systems account does not deny this, but nevertheless insists that the key perspective for normative evaluation of function is the current system rather than past selection. Regulation does not succeed by making parts function as they did in the past, it succeeds by making the system work well in present conditions.

This is only a provisional response; a detailed theory of what it is to have normative perspective is needed. The autonomy-based account at least takes some steps in this direction, and in this respect does more than the etiological theory. The etiological theory offers no explicit account of normative perspective, and the normative perspective that appears in the account is rather dubious. It is perhaps also worth briefly noting that the kind of normative perspective proposed by the autonomous systems account differs from a traditional consequentialist view in ethics in that it does not depend on the experience of pleasure and pain. It also does not appeal to a capacity for reflection, as with personal autonomy, though it does emphasize self-governance, albeit of a much simpler kind. The autonomous systems account does not aim to explain moral responsibility, or why certain entities should be the object of moral regard, so the type of normative perspective it proposes should be distinguished these kinds.

\(^4\) For a more detailed discussion see Christensen and Bickhard (2002).
5. Design and purpose

One reason why the autonomous systems account may seem less normative than the etiological account is because it does not appeal to design. Intuitions about design are a major buttress for the etiological account, and in particular the intuition that intricate functional structure of biological systems is a lot like the functional structure of artifacts. One of the key sources of normativity for artifacts (and their parts) is the intentions of the designer. An artifact is constructed so that the parts will interrelate in a way that conforms to the plan of the designer, and conforming to this plan is supposed to allow the artifact to perform its intended functions. By analogy we can regard biological organization as ‘virtual design’ (cf. Dawkins, 1986; Kitcher, 1993). In contrast, the autonomous systems account makes no appeal to design, and so is approaching the issue of normative function in biological organization from a very different direction.

Although the appeal to design helps give intuitive force to the etiological approach, it is not clear that this support is legitimate for the kind of reasons described in the previous section. That is, the analogy between biological systems and artifacts is questionable in exactly the ways that bear on normative function: ‘mother nature’ is not a real agent with real design intentions. In any case, though, even for artifacts, where there are real agents, we should not treat design intentions as the sole source of normativity, and perhaps not even the most fundamental source. An artifact can fail functionally even though it conforms to its design plan, and it may be functionally successful despite not conforming to the designer’s intentions. Valuational normativity must be more fundamental than design normativity because we need it to understand design itself: things do not work well just because they have been designed – designers try to design things that will work well.

Rather than base the normativity of artifacts on design we could base it on purposes (Franssen, 2006), treating design as just one source of purposes, with the users of artifacts being another. This
would help explain how an artifact might be functionally successful despite not performing according to its design, because normativity is also being conferred by the purposes of users. However a very broad understanding of purposes will be needed. Our interaction with artifacts is complex, especially in the case of artifacts that are themselves complex systems, such as buildings, aircraft, power stations, rail systems, computers, computer networks, etc. It can be difficult and sometimes impossible for the designers of such artifacts to understand in detail how the users will respond to the artifact, and the users themselves may not understand their interactions with the artifact particularly well, not least because tacit learning plays a major role. Moreover, we often adjust our purposes through experience with an artifact, both abandoning goals and discovering new ones. So the normativity of artifacts will not be well captured if it is thought of as entirely dependent on explicit psychological purposes. Although explicit purposes are undoubtedly an important source of normativity for artifacts, there also appears to be a form of normativity involving broader relations to human activity.

If purposes themselves are normatively constrained, such that we can have the wrong purposes, and through learning acquire better ones, then we need a deeper form of normativity. The autonomous systems account supplies one proposal for what this normativity might be. Artifacts are not themselves autonomous systems⁵, and only acquire normativity through their relations to users, who are autonomous systems. Admittedly human autonomy has rich psychological structure, but this psychological structure, rather being the fundamental source of normativity, is itself normatively constrained in virtue of being embedded in an autonomous system, which it helps to steer. The sources of normativity for artifacts and living systems are thus quite different: artifacts derive their normativity from their relations to the living systems that use them, whereas living systems have their normativity indigenously, in virtue of being autonomous systems.

⁵ In the future we may develop the ability to construct autonomous systems.
6. Reasons and persons

At this point it will help to revisit the question of what normativity is. On one view normativity is connected to reasons. Thus, ‘[a]pects of the world are normative in as much as they or their existence constitute reasons for persons, i.e. grounds which make certain beliefs, moods, emotions, intentions or actions appropriate or inappropriate’ (Raz, 1999, p. 354). This way of framing normativity is helpful because it highlights three key issues: (i) the perspectives for which things matter (persons, according to Raz), (ii) the nature of mattering (that which makes facts about the world relevant to entities with a normative perspective, these relevance relations constituting normative facts), and (ii) the mechanisms by which the entities with normative perspective respond to normative facts (which Raz associates with rationality).

Because it emphasizes personhood as the basis for normative perspective, and rationality as the mechanism by which persons respond to normative facts, this way of conceptualizing normativity may seem to favor non-naturalism.6 It at least confines normativity to the realm of rational agency, and thereby separates it from the broader natural world, even if we seek to treat persons and rationality naturalistically. This would be bad news for the autonomous systems approach because, although artifacts might derive normativity from their relationship to humans, much of the functional structure found in the living world is independent of humans and would not be normative.

[Table 1 about here]

However, the schematic structure of the conception of normativity given by Raz is congenial to the autonomous systems account, which proposes a similar structure involving more basic entities and

---

6 Olson (2009) argues that the recent emphasis on reasons has bolstered non-naturalism in meta-ethics, though not for the reasons suggested above. He argues that a reasons-based conception of normativity appears (wrongly) to make non-naturalism less metaphysically problematic than the Moorean version based on goodness.
mechanisms (table 1). The autonomous systems theorist must reject the idea that personhood and rationality are normatively fundamental, and propose instead that these are grounded in the more basic kind of normativity identified by the autonomous systems account. Indeed, two kinds of grounding are on offer: origins and constitution (figure 1). With regard to origins, the basic idea is to treat personhood as just a particular kind of agency, and more specifically as a cognitively sophisticated form of agency that has evolved from more basic non-cognitive forms of agency. The normativity of personhood is an elaboration of the normativity of these simpler forms of agency, with special features arising from the psychological attributes of personhood, but also with a great deal of continuity. With regard to constitution, persons are not just descended from autonomous agents, they are autonomous agents: a person is constituted as a certain kind of autonomous agent in the base sense of autonomy, and this makes an important contribution to the normativity of personhood. The normativity of reasons is in part the normativity of the functional organization that constitutes the autonomous system that is the person. Psychological mechanisms that respond to this normativity are complemented by biological regulatory mechanisms.

[Figure 1 about here]

Non-naturalists are likely to agree that there is some important story to be told about the relations described in 1a and 1b. However, they will restrict normativity to the far side of the arrows in each case. The idea will be that, although in the larger picture we are interested in knowing how cognitive agents have causally appeared in the world, and how they are causally instantiated, this is not the subject matter of normative theory as such, which is concerned only with the boxes the arrows point to. There are at least two kinds of response to a restrictionist view of this kind: (1) argue that structural parallels between the two boxes warrant extending normativity to encompass them both, and (2) argue that the restrictionist view will render its own subject matter incomplete and mysterious. The parallels described in table 1 are a starting point for an argument of the first
kind. Arguments for (2) can focus on the following two claims: (a) biological and psychological normativity is integrated in cognitive agents, and (b) psychological and biological mechanisms form a single normative response system (albeit imperfectly integrated) with numerous interdependencies, just as one would expect, given (a).

The following arguments provide some support for (a) and (b): (i) Psychological and biological mechanisms respond to the same normative facts. Thus, persons have reason to avoid consuming things that will make them ill. Decisions to avoid particular foods based on acquired knowledge and experience are a cognitive means for responding to this normative fact, whereas vomiting after ingestion is a biological regulatory mechanism for responding to the same fact. (ii) Biological mechanisms can respond to normative facts without the aid of psychological mechanisms. For instance, you eat food you believe is ok, but your body detects toxins and reacts with vomiting. (iii) Biological mechanisms can train cognitive mechanisms on which normative facts to recognize. You will for example learn to avoid foods that make you nauseous.

7. Applying the autonomous systems theory

Applying the autonomous systems theory of normative function to several examples will help make its structure clearer, and hopefully show its usefulness.

7.1. The remarkable re-wired ferrets

Experiments in ferrets have shown surprising functional plasticity in sensory processing areas of the brain. By deafferenting the auditory thalamus in ferrets at birth, Sharma et al. (2000) were able to induce retinal axons to innervate the medial geniculate nucleus (MGN), which is a relay to the ferret to primary auditory cortex (A1). In other words, visual input in these ‘rewired’ ferrets was directed in the first cortical area involved in auditory processing. Histological examination of the affected cortical area showed that it had taken on structural characteristics (orientation modules) similar to
primary visual cortex (V1) and unlike the typical organization of auditory cortex. Follow-up experiments reported in von Melchner (2000) addressed the question of how this highly abnormal area of primary sensory cortex was treated by downstream neural processing. A particularly intriguing question was whether, from the point of view of the ferret, input to the rewired sensory cortex would be treated as if it were visual or auditory. Given that no cortical areas downstream of A1 were directly affected by the intervention the most parsimonious prediction is that the ferrets will respond to stimuli processed through the re-wired A1 as if it is auditory.

Because only a known section of the retina was rerouted, with the rest of the retina connecting by the normal pathways to visual cortex, the issue could be tested by selectively presenting stimuli to the affected retinal areas. Restraining the ferrets in an apparatus ensured that a visual stimulus presented in external space would be processed by a restricted area of the retina. Using signals routed through normal sensory cortex the ferrets were trained to visit a reward spout on their left for an auditory stimulus and a reward spout on their right for a visual stimulus. The key question was, when presented with stimulus that was selectively processed by the modified retinal area, whether the ferrets would respond as if they had received an auditory stimulus or a visual stimulus. The results indicated that they treated it as visual: they went to the right reward spout. This result is striking because not only did the affected area of primary auditory cortex remodel itself for visual processing, somehow downstream cortical areas were able to detect that the information stream was visual rather than auditory, and exploit the information functionally in the control of behavior. Nevertheless, though dramatic, these findings are consistent with a wide range of evidence indicating high levels of plasticity in neural processing (see e.g. Elbert et al., 2001).

This example illustrates some of the limitations of assigning functions to parts without regard to the whole system. Humans and animals show a robust ability to recover from serious brain injury, often by constructing highly unusual functional circuitry. What will work well in the here-and-now may
be substantially different to what has worked in the past, and adaptive plasticity is a crucial mechanism that allows organisms to construct workable solutions in the here-and-now. It highlights the fact that what really matters, functionally, is that the current system work well.

Indeed, adaptive plasticity is one of the more important empirical phenomena that the autonomous systems account can help illuminate. There has been interest in the immune system for its role in distinguishing ‘self’ and ‘non-self’ (Tauber, 2010), but less appreciation that functional regulation in general poses questions about system identity and ‘better’ versus ‘worse’ normativity. Regulation alters system state in ways that, to be adaptive, must count as improvement for the system. It also acts to restrict what is incorporated into the system, and eject things that will have a negative effect. Cell membranes and the skin of multicellular organisms serve as regulated boundaries, whilst the regulation of material ingress via the mouth is especially complex and sophisticated, for obvious reasons. Sensory and motor systems provide more distal regulation of intake by means of approach and avoid behavior. Plasticity allows adjustment to local circumstances, but to be adaptive plasticity must be regulated so that it is shaped into functional forms that are beneficial for the system. The kind of dramatic neural plasticity seen in ferrets illustrates just how subtle and powerful the regulation can be. But behavioral learning is also a form of adaptive plasticity with profound effects. Almost all animals are capable of at least simple forms of learning, and many are capable of very complex forms (see e.g. Moore, 2004).

To develop a fundamental theoretical understanding of adaptive plasticity we need an account of the ontology of biological systems, and what counts as better or worse for them. Regulatory mechanisms depend on proxy information, e.g. a looming visual stimulus as a proxy for danger. The relation between the proxy signals and the underlying system conditions they regulate can be indirect and imperfect, but to be adaptive it must be the case that the regulatory mechanisms tend to have a beneficial effect. Thus, to understand the evolution of such mechanisms we need to
understand: (a) the proximal discriminations made by regulatory mechanism and the alterations they induce in system state, and (b) what the system actually is, and what actually counts as better or worse for it.

7.2. Reconstructing Joe

In the not-too-distant future these theoretical questions concerning normative function are likely to gain increasing practical relevance. Consider Joe, a fighter pilot of the late twenty-first century who has been involved in a collision during combat training in low earth orbit. He is extracted from his damaged craft and rushed by ambulance shuttle to an earth hospital. Joe’s main injury is that his right front cortex is smashed in from the anterior prefrontal cortex to the motor cortex. Fortunately, brain reconstruction can restore a high level of function, though Joe will never be quite the same as he was before the crash. His long-term declarative memories are largely intact, but a great deal of his motor skill is lost, as is much of the substrate for his higher order emotional and cognitive control. Because the reconstruction cannot exactly duplicate the organization of the lost neural tissue, and because Joe’s intact brain will reorganize during the reconstruction process, there will inevitably be significant functional discontinuity between Joe’s brain before the accident and after the reconstruction. Joe’s doctors cannot simply recreate the functional system as it was before the accident, they have to create new functional structure that aims as best possible for the continuity of Joe-the-person in his new circumstances. To know how to intervene in Joe’s brain his doctors will need to develop a rich understanding of Joe-the-person.

Joe’s brain is unique; it has been molded by his genetics, development and idiosyncratic learning experience. Using scaffolded neurogenesis the doctors will begin to reconstruct the basic structures of the right frontal cortex, and induce the major projections to and from other cortical areas. But the generic neuroanatomical templates must be adapted to the specific structure of Joe’s brain, and the fine-grained structure of the rebuilt neural systems must mesh well with the rest of his brain. As a
fighter pilot Joe had extremely well developed higher cognitive control: excellent control of attention, high ability to maintain spatial and task awareness, excellent task management ability, and so on. A great deal of this cognitive control was provided by the brain areas now destroyed, and brain areas that were previously under complex patterns of regulation from the right frontal cortex are no longer experiencing this regulation, with the result that they will tend to disorganize and organize for other functions. Unless Joe’s doctors take steps to prevent it Joe’s brain may start to form bizarre and dysfunctional patterns during the course of the reconstruction.

Thus, to achieve functionally successful reconstruction Joe’s doctors will need to do more than regenerate the basic neuroanatomy of the right frontal cortex, they will need to exert precisely structured influence on the re-growing neural systems to shape them in relation to higher cognitive functions and the overall organization of Joe’s brain, and Joe-the-person. Figure 2 sketches a systems-based ontology for Joe, together with some of the kinds of norms that come into play with various aspects of the ontology. According to this ontology Joe-the-person is an individual with a particular kind of lifestyle, namely being a fighter pilot. In turn, as a fighter pilot Joe is at a more basic level a cognitive agent, and at an even more basic level he is an autonomous system. These are hierarchically structured forms of organization, each of which impose normative constraints. As an autonomous system Joe is subject to the very basic norms of existence for an autonomous system. As a cognitive agent Joe is subject to general norms for agency and cognition. Many of the core functional norms that apply in the reconstruction of Joe’s brain come from here: to be a competent cognitive agent Joe will need amongst other things functional working memory, reasoning and higher order emotional regulation. As a fighter pilot norms for this skill domain apply, with excellent visuo-spatial working memory and fine motor control being some of the more important. If he is to return to his previous life the doctors will need to rebuild these capacities. There are also norms specific to Joe: Joe-the-person has a particular history and personality, specific friends and family, and particular cognitive skills acquired through idiosyncratic learning.
shaped by his particular cognitive strengths and weaknesses. To return to his previous life he needs to maintain and continue to develop his individual cognitive and social style, and his personal relations. He must not only be good at being a fighter pilot, but also good at being Joe.

[Figure 2 about here]

It may be that it is not feasible to re-make Joe in a way that allows him to return to his old life, or at least not obvious that this is the right thing to do. The reconstruction of the basic neuroanatomy is likely to take many months, followed by a much longer period of therapy and training designed to induce the formation of the appropriate fine-grained functional organization, much as with the post-acute phase of stroke rehabilitation now. In the latter part of this century brain regenerative techniques are much more precise and effective than they are currently, involving carefully targeted cognitive and behavioral therapies complemented with direct neural therapies which include an array of implants that provide electrical stimulation, targeted delivery of growth factors, and structural scaffolding. Even so, after such a massive injury it will take years for Joe years to regain his former elite abilities, by which time the nature of his job will have changed and his former companions moved on. Joe’s left, language-oriented hemisphere is intact, and he has always been witty and verbally gifted. He read extensively, and had talked with his wife and friends about becoming a writer after retiring from the military.

This presents Joe and his doctors with alternatives that become increasingly distinct as the rehabilitation process progresses. If the aim is to make him a fighter pilot then neurocognitive rehabilitation should focus on developing a quite different set of cognitive abilities than those required to be a writer. The training needed to develop either suite of abilities to an advanced level is intensive and protracted. Training that aims to help him become a writer will amongst other things emphasize higher order emotional and social cognition, rather than visuo-spatial cognition,
motor control, and task management under time pressure. Joe is allowed to regain consciousness less than a week after the accident, but with heavy stabilization to damp disorganized brain activity in response to the injury, and traumatic psychological response to his situation. With his right frontal cortex destroyed his cognitive and emotional regulation is substantially impaired, and he will not be competent to make complex choices about his future until much later in the reconstruction. His wife and family are consulted extensively, and the doctors delay putting a particular functional orientation in the reconstruction until Joe’s basic decision making ability improves.

The holistic structure of the normative constraints in this situation can be characterized in terms of a notional decision cycle (Figure 3). First, there is an assessment to determine whether the system is or can be an autonomous system. In the case of Joe, the initial question is whether he will survive. The next step in the cycle is to determine, given the current system capacities, and intervention capacities, the best available state for the system. Here goals for the reconstruction are determined: re-establishing Joe as a competent cognitive agent, remaking Joe-the-person, giving Joe the abilities he will need for a suitable career. These goals guide the detailed neural interventions, and as the goals become increasingly specific Joe’s doctors can work out in increasingly precise ways how particular parts of Joe’s brain should be functioning during and after the reconstruction. To adapt to changing low and high order information Joe’s doctors will go through many decision cycles with a similar structure to Figure 3, though the emphasis will shift from whether he can survive to how he should be.

[Figure 3 about here]

The dominant role of the second step (determining the best available system state) in shaping the third step (normative role assignment for system elements) reflects the holistic structure of the
normativity of functional value, which has its source in the fact that the system as a whole constitutes the key normative perspective. Parts have no particular value or normative role independently, and history exerts no direct normative influence. History certainly exerts a powerful indirect influence: Joe has an extensive legacy from his life before the accident, and this legacy – including his intact memories and knowledge, existing personal relations, home and possessions, for example – tends to be more supportive of some ways of being compared with others. It in particular tends to be more supportive of his former way of life relative to alternatives. But Joe’s past exerts normative influence only by shaping the ways of being available for Joe now, and the legacy from his past may not be decisive in favoring Joe’s previous way of life, especially in the drastically altered circumstances he is currently faced with. Thus, to reiterate, the crucial normative perspective is Joe-now, and what possibilities there are for Joe-now.

Figure 3 goes beyond the account given in section 4 by incorporating a form of normative role assignment. The account of section 4 only addresses value. In the human context, where we are considering direct intervention to modify function, the cognitive apparatus at work generates goals and plans, and role assignment. It may be that more limited forms of role assignment can be characterized in purely biological cases, but I will set aside this issue. The basic account of normative function described in section 4 is agnostic with regard to role assignment.

7.3. Matilda’s slippery relation to the truth

A very different kind of example may help to further clarify the system-relative nature of functional norms. Hattiangadi (2006) argues in favor of naturalism for meaning. She accepts the argument that if meaning has prescriptive normativity then it will violate Humean limitations on naturalist explanation, and that we will therefore need to take a non-naturalist approach to meaning. Her aim is to show that the normativity of meaning is descriptive rather than prescriptive⁷, and that

---

⁷ She doesn’t distinguish valuational from prescriptive normativity.
naturalism is consequently safe. The gist of her argument that meaning normativity is descriptive is that, although meaning has correctness conditions – it is correct to apply the term ‘horse’ to X if and only if X is a horse – meaning is not prescriptive inasmuch as speakers are not obliged to use terms correctly.

Hattiangadi distinguishes hypothetical means/ends prescriptions, which specify conditional relations of the kind ‘If you want to get to the airport on time, take a taxi’ from categorical prescriptions, which are not conditional on ends. She claims that hypothetical means/ends prescriptions do not pose a difficulty for naturalism because they are not really prescriptive. Even the combination of a hypothetical means/ends prescription and the appropriate end is not prescriptive, because it might be the case that the end should be abandoned. Thus, even if it is true that, if you want to get to the airport on time, you should take a taxi, and it is also true that you want to get to the airport on time, it still does not follow you should take a taxi, because there might be reasons that make it better for you to abandon the goal of getting to the airport on time. Perhaps, rather than going on holiday, you should stay to look after your sick parent. Hattiangadi claims that the normativity of meaning is hypothetical and consequently merely descriptive. Accordingly, Matilda, who tells terrible lies, should use her words correctly if she wants to tell the truth, but she is not obliged to tell the truth.

However, even if Hattiangadi successfully shows that meaning is not prescriptive her argument leaves naturalism in an uncomfortable situation. After all, surely there will come a point where Matilda ought to do something. Either there is not, in which case it looks like there is nothing anyone ought to do, or there is, and it looks like naturalism will be in trouble at that point. Not perhaps in virtue of meaning normativity, but possibly in virtue of some norm of practical reason. Of course, the standard Humean answer is that what Matilda ought to do is determined by applying the maximizing principle to her total set of desires. This is naturalistically safe, supposedly, because
no objective norms are appealed to, only Matilda’s desires. However the Humean answer has some unattractive features. The maximizing principle itself appears to be a normative principle, so there is the threat of inconsistency (Wallace, 2008), but the approach also rules out the possibility that Matilda could have the wrong desires. It will fault her if her desires are inconsistent, but it does not allow that Matilda might have a consistent set of desires that are misguided.

It seems on the face of it quite possible that Matilda might have a misguided set of desires, and the autonomous systems view provides an account of what this might amount to. As a person she is a particular kind of autonomous system with a complex normative structure something like that depicted in Figure 2. Her desires are proxies for what is actually good for her, and they can be misaligned with what is good for her, as well as with each other. This idea runs contrary to the Humean separation of fact and value, but the benefits are substantial. By not making purposes normatively primitive we get to understand the functional relations between psychological structures and the systems they steer. This helps us understand how these psychological structures have evolved, and how they can be improved in the here-and-now. Furthermore, the account of valuational normativity draws on naturalistically unproblematic resources.

The autonomous systems account challenges Hattiangadi’s framing assumptions, but on the other hand it supports her idea that prescriptive normativity should not be directly associated with meaning. According to the autonomous systems account such normativity can only be assigned after taking into account the larger system. Within this framework meaning norms might be interpreted as constitutive norms pertaining to a certain kind of ‘game,’ specifically the ‘game’ of rational cognition and communication. Matilda-the-person is not obliged to play this game on all occasions, but as a cognitive agent she is structurally committed to being rational at least some of the time. That is, if she is not rational at least some of the time she will cease to be a cognitive agent. The constitutive norms of meaning gain valuational and prescriptive normative force for her
because of this structural commitment: in thought and language it is sometimes good for her to use concepts and words correctly. In effect, the constitutive norms of meaning are also partly constitutive for her as a cognitive agent, and because she has normative perspective her constitutive norms have evaluative normativity (for her). But her agency does not depend on perfect conformance to the constitutive norms of meaning, and whether she should adhere to such norms on any given occasion is a function of a complex array of personal and situational factors. Taken individually and without regard to context, meaning norms are not prescriptive.

8. Conclusion: natural sources of normativity

Naturalist approaches to function, cognition and agency may have hobbled themselves unnecessarily by restricting themselves to ‘descriptive’ normativity. The putative normativity is thin at best, and without an account of valuational normativity we are left with an incomplete understanding of key phenomena like regulation and adaptive plasticity. Conversely, approaches that ground normativity in high level features of human agency, such as personhood or purposes, also leave us with an incomplete and somewhat mysterious picture. The structures and capacities that support high level agency are themselves, arguably, constrained by broader forms of functional normativity. A naturalist approach that tackles evaluative normativity head-on, rather than skirting it, can provide a more coherent and informative picture.

The discussion here has aimed to make these claims plausible, but a detailed theory will need to address a number of difficult issues. The ontological analysis of autonomous systems must be defended, and a detailed argument linking the features of autonomy to normative perspective is required, together with a more general account of normative perspective. It will also be important to specify more closely how different global system states are to be evaluated as better or worse; in other words an account of flourishing is needed, and it must avoid succumbing to circularity or stipulation. The extension of the ontology of autonomous systems to encompass personhood, along
the lines suggested in Figure 1, will confront many highly contentious issues. It is not unreasonable to worry that the approach may not be able to satisfactorily carry through on these tasks, but equally, it seems worth trying. There is a prima facie basis for thinking that the sources of normativity, at least of the kind considered here, are natural systems amenable to broadly scientific understanding.

Acknowledgements

Preparation of this manuscript was supported by the Australian Research Council grant ‘From Neuron to Self: Human Nature and the New Cognitive Sciences’. The ideas presented here build on earlier collaborative work conducted with Cliff Hooker and Mark Bickhard.
References


Tables and Figures

**Table 1: Structural parallels between reasons and autonomous systems conceptions of normativity**

<table>
<thead>
<tr>
<th>Normative perspective</th>
<th>Basis of mattering</th>
<th>Mechanism for responding to normative facts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasons conception</td>
<td>Persons</td>
<td>Relevance to the person</td>
</tr>
<tr>
<td>Autonomous systems conception</td>
<td>Autonomous systems</td>
<td>Relevance to the autonomous system</td>
</tr>
</tbody>
</table>


Figure 1: Two kinds of naturalist grounding for the normativity of personhood.
Figure 2: System-based specification of Joe's ontology, together with associated norms.
Figure 3: A notional decision cycle for determining how to intervene to improve function.