

Nature, agency, and the nature of agency

Kenneth Walden

Abstract: I examine skeptical arguments about the constitutive nature of agency, with special attention to those of Elijah Millgram. I suggest that these arguments lead us not to the conclusion that agency has no such nature, but that it is an essentially contested kind in the same way that art is. I argue that this undermines traditional forms of constitutivism in metaethics but opens the door to a different way of pursuing the same program. Finally, I take issue with Millgram's solution to the problem of "logical aliens" and suggest an alternative based my analogy with art.

Keywords: agency; practical reason; constitutivism; art; Elijah Millgram.

1. *Nature and agency*

Live in agreement with nature. So says Cleanthes the Stoic. This advice has been hard for philosophers to shake: from Aristotle's axiom that the human "is by nature a social being" to Bentham's "two sovereign masters, pain and pleasure" that "nature has placed us under," few ethical theories avoid turning their spade on a thesis about human nature. But these views face two obvious problems. *Which* "nature" am I supposed to live in agreement with? I am an organism, a vertebrate, a mammal, a primate, a human being, a rational animal, a social animal. I live in a world governed by laws of physics, chemistry, biology, even zoology. Which of these am I to live in agreement with? And *why*, moreover, should I so live? If I want to be a hermit, what reason does my membership in a social species give me to live otherwise?

These problems are especially acute when the "nature" we select has a tenuous connection to practical reason. I am indeed made mostly of carbon, but it is not in the guise of carbon-based lifeform that I engage in practical reasoning, nor does this fact do much to structure that reasoning. Unfortunately, the same seems to be true of many of the natures that the followers of Cleanthes are likely to turn to: the things that set *homo sapiens* apart from plants and even other primates. I needn't represent myself as, e.g., a social

creature when undertaking practical reason – or at least it's not obvious that I must – and for this reason the question of why I should care about being one remains open.¹

One version of Cleanthes's dictum seems designed to overcome just this problem. Constitutivists say that we should live in agreement with our nature *as agents* – that the constitutive features of the sort of agency we are capable of correspond to universally and unconditionally authoritative normative principles. We ought to so live, they go on to say, because we cannot really do otherwise. Or, more precisely, the demands of agency are normatively inescapable because it is only from the point of view of agency that normative questions of how to live arise at all, and this is just the point of view from which our agency is presupposed. So it may be metaphysically possible for us to be otherwise than agents, but insofar as our agency is a presupposition of our practical questions this is a practical impossibility. This connection gives our nature as agents an apparent normative significance lacked by our nature as carbon-based lifeforms.²

But precisely because constitutivism promises such a graceful escape from our initial challenges, it faces a different problem. Is there anything that is actually ruled in or out by the requirement to live in agreement with one's nature as an agent? How could this requirement be anything but vacuous? Of course, philosophers have claimed to have extracted interesting requirements from the nature of agency as such. Korsgaard (2009) claims that adherence to the Hypothetical Imperative and the Formula of Universal Law are such conditions. Velleman (2009) claims that aiming at self-understanding is one. Katsafanas (2013) claims that structuring our activities so we are constantly overcoming challenges is one. But these claims have been greeted by accusations of equivocation. The constitutivist employs a pristinely generic conception of agency when arguing that the conditions of agency are inescapable normative standards, the charge goes, but then subtly but illicitly switches to consideration of a particular *subkind* of agency when deriving these particular conditions – a kind that, unlike generic agency, is optional.³

Charges of equivocation will necessarily concern individual constitutivist arguments, and this limits their effect. We may be able to diagnose equivocation in the arguments of individual constitutivists, but this will leave it

¹ Though see Foot 2003: 52-65 for an attempt to close it.

² See, e.g., Ferrero 2009, Velleman 2009: 138-139, Silverstein 2015.

³ Tiffany 2012 puts the challenge in these terms. The multifaceted "shmagency" objection pressed in Enoch 2006 and Enoch 2011 can also be read in this spirit.

open that there is some better version of the argument over the horizon that succeeds where these fail – an argument that latches on to the *true* constitutive nature of agency and finds something genuinely interesting there. We might wonder whether there is any way we can foreclose this possibility and show that constitutive arguments are destined to equivocation or triviality. One view that would seem to do the trick is what Elijah Millgram (2010: 90) calls “pluralism about action”. This is the view that “‘action’ (or ‘agent’, the other side of the conceptual coin) is a family resemblance concept”, just as Wittgenstein argued *game* is: “there are no features that all games share, and so, even if people necessarily play only games, because there are no substantive and true claims to the effect that games are such-and-such, you cannot argue successfully that, if one is a player, one inevitably engages in activities that are such-and-such”.

Establishing a view like this is tricky. Millgram’s principal evidence that *agency* and *action* are family resemblance concepts seems to be that the interesting claims about what distinguishes agency turn out to admit significant exceptions. This includes both the constitutive theses we saw before as well as more anodyne claims (e.g. that action involves progress toward end points and that it can be decomposed into smaller units).⁴ But these arguments are not dispositive. Why, a hopeful constitutivist may ask, is this evidence for the claim that *agency* (or *action*) is a family resemblance concept, and not just an isolated failure to discover the true constitutive nature? To this Millgram could respond – though he doesn’t say quite this – that these failures add up to a sort of pessimistic induction. None of our attempts at analysis have succeeded so far, and that gives us reason to doubt that *any* analysis forthcoming will. I don’t think this version of the argument can work, though. For it only seems appropriate if we think that attempts to characterize the constitutive nature of agency are the sort of things amenable to induction. But that possibility seems ruled out by the very claim in question – that *agency* is a family resemblance concept. For in that instance the class of agents would be too heterogeneous for an inductive argument to be persuasive. We don’t have the same problem with the concept *game* because it is an artefactual concept, and we can be more certain that its surface heterogeneity does not belie some deep, hidden, unified nature waiting to be discovered by the right sort of philosopher. Whether *agency* is like this, however, is precisely what’s at issue.

This awkward position makes it difficult to transform pessimism about constitutivism premised on the failures of particular versions of the thesis into a general

⁴ Here also see Millgram 2015 and Millgram 2016.

argument that agency is too disjointed or heterogeneous for any version of the program to succeed.

2. *Agency's nature*

For this reason I doubt that pessimistic arguments like these can move us very far beyond the game of whack-a-mole we were playing before. Fortunately, Millgram does have another, more far-reaching argument. He notes that “the central competing accounts of action amount to different models of process control”, which suggests a useful analogy “between distinctive programming styles (along with the programming languages meant to enforce them), and the models of process control that distinguish competing accounts of action”. This analogy gives us reason to “think twice as to whether we can treat any such model as a theory of (all) action [since] modularized programming typical of function-based languages was the product of something of a revolution in the field”. He continues:

Just as structured programming was an achievement, so analogously structured action is evidently also an achievement. Because it is an achievement, it is not what actions turn out to be, willy-nilly. To the extent that the metaphysics of action has to do with the control structures that organize activity, the analogy makes it clear that our real question cannot be: What are actions (as a matter of metaphysical necessity)? And so we have identified a second assumption of most previous work in action theory: that the question of what actions are asks about what they already are, as a matter of standing fact. [...] It would be crazy for computer scientists to argue about what style of programming is metaphysically necessary (as though *it* were a matter of standing and immutable fact); I suggest that it is likewise an unpromising avenue of approach to their subject when philosophers argue over what control structures are metaphysically essential to actions (as though *that* were a matter of standing and immutable fact). The structure and composition of actions is evidently an engineering problem, one which can take various and novel solutions: the real question is not what actions *are*, but what sorts of control structures for action we can design and implement, with an eye to whatever benefits are to be had from them. (2010: 93-95)

Computer programs constitute a kind whose nature is gradually *invented* as a response to design problems, and this is the way they must be. Someone who thought that it was part of the constitutive nature of computer programs to be written in COBOL wouldn't just be wrong; they would be in the grip of a fundamental misunderstanding about the sort thing a computer program is. Millgram's analogy therefore suggests a less circumstantial argument than the one we just saw. Agency and action may be like computer programs in-

so far as they lack a fixed constitutive nature *of necessity*. This suggestion is different from the family resemblance claim. In saying that *game* is a family resemblance concept, we mean that the class of games in fact lacks the unity or similarity necessary for there to be anything constitutive of gamehood, that there *happens to be* nothing unifying the class. The present claim, suggested by the programming analogy, is that agency *must be* the sort of thing that lacks an antecedent constitutive nature – that if we assume it has a constitutive nature we are fundamentally misunderstanding the kind of thing it is. To put the point in a way that courts paradox: agency is a kind whose constitutive nature it is to lack a constitutive nature.

It's not completely clear what this thesis amounts to. The best way to investigate it is by considering how we might argue for it. I want to consider a few arguments here. The last of these comes from Millgram himself, but he has some notable predecessors.

One of the earliest accounts of agency that holds the potential for entailing this thesis comes from Kant. Kant is interested not in agency *sans phrase*, but in rational agency – the agency of those creatures with a rational will. So in understanding how Kant conceives of agency, we should consider how he understands the faculty of reason. Kant's rationalist predecessors understood reason as a faculty for discerning a certain species of truth, and thus very roughly of a piece with a capacity like vision. In this spirit, they supposed that reason was responsible for our knowledge of arithmetical and geometrical truths, and this gave them hope that the principles endemic to philosophy might be gleaned in the same fashion. One of the principal conclusions of the *Critique of Pure Reason* is that this picture of reason is both a mistake and the source of mischief. For one, the truths of mathematics are constitutive not of the operation of reason but of the form of inner and outer intuition. More importantly, reason is not a faculty for detecting a reality, but a *critical* faculty – a faculty characterized not by its ability to produce answers, but by its tendency to raise questions, to reflect on our judgments, and challenge their bona fides. Thus Kant (1781/1787, Eng. tr. 1998: A795/B823) says that “the greatest and perhaps only utility of all philosophy of pure reason is thus only negative, namely that it does not serve for expansion, as an organon, but rather, as a discipline, serves for the determination of boundaries, and instead of discovering truth it has only the silent merit of guarding against errors”.

Now Kant's discussion in these passages is almost exclusively about reason in its theoretical use, but it nonetheless seems fair to conclude that the kind of agency made possible by the critical faculty of reason will be rather different from that associated with a productive faculty – an “organon” as he

calls it. Kant spells out some of these differences in his entry in the uniquely Enlightenment genre of conjectural history. Thus his focus in the *Conjectural Beginning of Human History* is the difference between creatures controlled by instinct and those that have the power of reason.

In the beginning the newcomer must have been led by instinct alone, this voice of God which all animals must obey. Instinct allowed him certain things for nourishment and forbade him others. [...] As long as the inexperienced human being obeyed this call of nature, all was well with him. (1786, Eng. tr. 2006: 8:111)

Instinct, so described, seems to be an internal motive for living in accord with a one's nature. The shark is driven by instinct to eat seals, and that is all well and good because part of what it is to be a shark is to eat seals. Over time, natural selection will keep instinct and the natural needs of a creature from drifting too far apart.

The introduction of reason changes everything. "What occasioned the desertion of the natural urges may have been a trifle, but the result of the first experiment, that is, becoming conscious of one's reason as a faculty that can extend itself beyond the boundaries to which all animals are confined, was very important and decisive for the way of life". He continues:

This was sufficient to give reason occasion to do injury to the voice of nature and, despite its protest, make the first experiment in free choice, an experiment which, as the first one, probably did not turn out as planned. However insignificant the harm done may have been, it sufficed to open the human being's eyes. He discovered in himself a capacity to choose a way of life for himself and not, as other animals, to be bound to a single one. The momentary delight caused by his noticing this advantage must have been followed by anxiety and fear as to how he, having not yet known anything according to its hidden traits and remote effects, should proceed with his newly discovered ability. He stood at the edge of an abyss, as it were. For whereas instinct had hitherto directed him to individual objects of his desire, an infinity of such objects now opened itself up to him, from among which he did not yet know how to choose. Yet once he had had a taste of this state of freedom it was impossible for him to return to the state of servitude (under the rule of instinct). (8:112)

Here we see the practical implications of reason understood as a critical faculty. If reason were an "organon" performing a particular function, like putting us onto nutritious things or carrying us from inborn premises to their conclusions, then it would just be another instinct – a more sophisticated instinct, surely, but functionally on par with the instincts of animals. But reason is not like this. By putting distance between us and our instincts, it frees us from "servitude" and puts us on the "edge of an abyss". This makes rational agency look quite different from the capacities for activity exhibited by other animals. As

Allen Wood (1999: 199) explains it, “reason is precisely our capacity for an indeterminate mode of life, one that is open-ended and self-devised, in contrast with the life of other animals, which is fixed for them by instinct”. This makes rational agency very different from the forms of agency exhibited by other creatures.

This contrast may lead us to think that Kant held that rational agency has no constitutive nature (i.e. that there are no interesting, non-trivial properties that human agency necessarily possesses). But that doesn't quite follow. It's true that rational agency doesn't have the kind of constitutive nature that the agency of a creature guided by instinct does. For these creatures, what is constitutive of their agency is a kind of fit between their nature as an organism and the instincts that direct them. Insofar as the squirrel's instincts drive it to forage, nest, and breed in the ways distinctive of squirrels, its squirrel agency is in good order. And this means that we will be able to identify properties relevant to the kind squirrel agency through a little zoology. Because rational agency is not reducible to instinct, nothing like this will be true for it, but that doesn't mean it must lack a constitutive nature altogether. Indeed, Kant's very characterization of rational agency relies on its possessing such a nature. Being regulated by a critical faculty like the one Kant describes in the first *Critique* rather than instinct is itself constitutive of rational agency, and thus part of its nature. This is a very different sort of nature from that possessed by instinctual creatures. It is not reflected in any *definite* quality – like a squirrel's drive to hide nuts or the sea turtle's impulse to paddle – but in the quality of *not* being so determined. I'll return to the question of what to make of this difference later, but for now I think we can conclude that Kant's account of rational agency, interesting as it is, does not support our thesis.

A more radical version of this kind of claim can be found in Sartre. Unlike knives and tigers, whose nature is fixed “in advance” of their existence by the kind of thing they are – by their essence – human beings are distinguished by *not* being so determined. Sartre concedes that just like knives and tigers we do have a “facticity”: we have masses and histories and causal powers. But unlike these other beasts, what we are is undetermined by these things. There is no gap, so to speak, between the knife's facticity and its nature: the knife's mass, length, sharpness, history, and so on determine what it is. But there is a gap between my facticity and my nature. All my analogous qualities still leave the question open what kind of person I am. We experience this shortfall from the first-personal perspective as “transcendence”: the experience of thinking that I can know everything about myself and the world around me and still wonder how to live. On the contrary, my nature can only be settled by my choices and how I create meaning in the world through my projects. (Sartre 1943, Eng. tr. 1956: 56-85)

We can easily rejigger this existentialist jargon to get a claim about human agency. The claim would be that unlike other kinds, whose nature is determined by their “facticity” – by constitutive features like being designed to cut, having a certain causal history – agency is precisely that kind which is *not* determined by such features. To be a member of this kind is, so to speak, to constitute oneself, not to hew to some antecedent standard. Set aside the question of whether this is at all plausible. If this claim about agency is right, does it follow that agency must lack a constitutive nature?

It does not, and for much the same reason that Kant’s claim did not. For the existentialist theory of agency begins by identifying a constitutive feature: the quality of being undetermined by one’s facticity. If x is the sort of thing whose entire nature is determined by its facticity, then x is perforce not an agent. And just as in Kant’s case, this not a trivial feature of agency discoverable through conceptual analysis. It is a substantive thesis we learn through phenomenological investigation (Sartre 1943, Eng. tr. 1956: 70). Our Sartrean account of agency ends up in a place much like Kant’s: it is strictly incompatible with the thesis that agency must have no constitutive nature even though the constitutive nature it attributes to agency is very different from the sort we might attribute to paradigmatic kinds.

A final account of agency begins in a rather different place. This one is Millgram’s. He starts with a distinctive orientation toward the forces that shape the nature of living things. He says that we can understand them as a solution to the problem of producing species suitable to available ecological niches. From this orientation we can see a typology of different “solutions” to this problem.

Weedy species are the ones that invade niche after ecological niche, and because they travel from one to the other, they’re not necessarily particularly tailored to any of them. Specialized species, on the other hand, are often fitted to their places in an ecology with a memorable and jewel-like precision. (2016: 55)

There are drawbacks to both weedy and specialized solutions. Weedy species are wasteful, and specialized ones are fragile. A designer might therefore have the idea of “build[ing] a species that was *both* weedy and specialized, that was in fact weedy *by* being specialized”. Millgram discusses two examples of this kind of solution. The first he calls Piltdown Man.

Piltdown Men are born, identify the environment they’re in, load a program appropriate to that environment from a database of available strategies, and run the program until they die. Piltdown Man can occupy what would intuitively look like many different locations in many different ecologies, and not the way that mints do: on the seashore, they might be fishermen; in the mountains, they might be yak herders; in the cities, they might be merchants; on the plains, they might be farm-

ers. However Piltdown Man can only occupy relatively stable niches because those software libraries have to come from somewhere; and if we don't have a Kubrickian monolith hand-coding them, they will be produced by a process that is either natural selection or (like meme selection) resembles natural selection in being, by human standards, slow.

According to Millgram, philosophical conventional wisdom has it that we are Piltdown Men. Most philosophers suppose that even though a person's nature as an agent is not determined in her evolutionary past – as may be the case for a squirrel – this nature is laid down soon after the individual emerges and she loads a “program” that stays with her throughout her life. What program a person runs is a contingent matter, but once loaded it has, for that person, an unshakeable normative authority. It is this picture, Millgram says, that makes instrumentalism about practical reason look like the default view. We think of our ends as aspects of this program and thus as non-negotiable inputs, and so see practical reasoning reduced to the technical problem of how to achieve those ends.

The trouble with all this, Millgram goes on to say, is that we are not Piltdown Men (indeed, noone was). We are *serial hyperspecializers*. We are not shackled to a single program for our whole lives but can exchange one program for another when we find our results are less than satisfactory. Nor do we need to rely on a “Kubrickian monolith” to install something out of the factory catalog. Instead, we can reprogram ourselves on the fly to meet the exacting needs of exacting niches. We can make ourselves a “VLSI engineer, comics inker, Cobra gunner, French professor specializing in eighteenth century poetry, adventure travel agent, director of cinematography,...”. As a design solution, serial hyperspecialization enjoys important advantages over Piltdown Man. It is less wasteful, since we needn't dispose of individual organisms whenever their program becomes sub-optimal, and the possibility of bespoke self-programming makes more exacting ecological fits possible. Once we recognize that we are serial hyperspecializers, the appeal of instrumentalism disappears. We realize that practical reason is not just the activity of carrying out a pre-loaded program, but also the activity of deciding which program to load.

If Millgram is right that we are serial hyperspecializers, then our agency lacks a constitutive nature in one important sense. It is not a matter of carrying out a particular pre-loaded “program” fitted to a single ecological niche, the way that the agency of migratory birds or Piltdown Men are. But like Kant and Sartre, this thesis itself attributes a different sort of constitutive nature to human agency: the claim that we are serial hyperspecializers suggests that the

capacity to “reprogram [oneself] on the fly” is itself constitutive of the special agency we possess.⁵ So it too ends up incompatible with our target thesis.⁶

3. *Agency as a contested kind*

These three accounts of human agency emerge from very different places. Kant is doing rational anthropology. Sartre is doing phenomenology. Millgram is doing evolutionary design analysis. But they all come to a similar conclusion: that human agency doesn't have the sort of constitutive nature that squirrels and knives do, and yet this very lack is attributable to a distinctive feature of agency that we might as well call agency's constitutive nature.

They also exemplify an obstacle in making the kind of ambitious argument I suggested. If we want to show that agency necessarily lacks a constitutive nature, we will probably end up needing to characterize agency in some way and then argue that the lack of a constitutive nature necessarily follows from that characterization. But then this very characterization ends up ascribing what looks very much like a constitutive nature to agency, and so the argument falls into a kind pragmatic paradox.⁷ In this respect the argument resembles our attempt at a pessimistic induction from before: in order to secure the necessary generality we need to make a claim at odds with our ultimate conclusion. This is a real pickle: I don't see how we could argue that agency is *necessarily* the sort of thing that lacks a constitutive nature without relying on an argument more or less like this, and so I don't see how we can reach such a conclusion without paradox.

Now this doesn't mean that the claim that agency lacks a constitutive nature is false. But I do think it dims the prospects for a decisive argument for the claim.

⁵ As it happens, Millgram makes concrete suggestions about how serial hyperspecializers reason about this programming. For example, he says that they rely on pain and pleasure in deliberating about their high-level ends and goals (2016: 61). This may be true for almost all actual human beings, but I don't see why it is *necessarily* true that pain and pleasure play this regulatory role (unless we define pain and pleasure in such a way to make the claim trivial). I can imagine a potential sadist deliberating about whether to load a program that assigns a rather different role to pleasure and pain. So if Millgram means to suggest that this reliance on pleasure and pain is constitutive of the agency of serial hyperspecializers, I think he is better off in a position that acknowledges only serial hyperspecialization itself as constitutive of human agency.

⁶ Millgram seems to recognize this. In a discussion of Aristotelian categoricals, he says that the only ones that appear stably true of human beings “amount to a description of the ways in which the species form is plastic” or, “the most important Aristotelian categorical about *our* species is that the Aristotelian categoricals true of it change from decade to decade.” (2016: 75, n. 35) Of course for the categorical Millgram mentions not to be self-contradictory, we must understand it as saying that the Aristotelian categoricals true of human beings change from decade to decade, *except* this very categorical.

⁷ Here compare (Millgram 2010: 95).

This leaves us with three stable positions on the question of whether agency has a constitutive nature. We could hold that agency has a constitutive nature in the way that knives and squirrels do, that some relatively definite qualities (like making plans or aiming for self-understanding) are constitutive of agency. We could be tentatively pessimistic that agency has any constitutive nature. Or we could think that agency has the very special sort of constitutive nature suggested by the accounts I just surveyed. The last of these is the most interesting, I think, and in what remains I want to flesh it out.

What is so special about the constitutive features that Kant, Sartre, and Millgram attribute to human agency? The obvious observation is that they attribute something negative to agency. They differ from a foil – instinctual animals, artifacts, Piltdown Men – by lacking some factor (the force of instinct, etc.) that would otherwise determine how the creature lived. In place of this factor, we have the suggestion that agents can *self-determine* in these same respects – that this determination is to be accomplished by the agent herself. It is a platitude that agency is a capacity for self-determination, but what distinguishes these views is that they hold that the parameters within which this self-determination is undertaken are *also* to be determined by the agent.

The crucial question is: how are agents to make this more radical kind of determination? How do they decide which “program” to load? How do they come up with an answer to the question put to them by reason? How do they transition from anguish to action? These questions impose an important constraint on potential answers. Let’s stick to Millgram’s way of talking, for simplicity’s sake. If we were Piltdown Men or squirrels, then Mother Nature or a Kubrickian monolith could program us before we gain consciousness and we could go about our lives without any awareness of that we are running a program that adapts us to a particular niche. But because we are serial hyperspecializers, we program ourselves, and that requires representing the program we are loading *as* a program.

Only certain things can be represented in this way. I can decide to program myself as a milkman or mathematician or even a Cobra gunner because these are all ways a person can be. But I cannot realistically decide to program myself as a Mayan Space Milkshake, since “Mayan Space Milkshake” does not denote anything. (More precisely: I cannot do anything that would make the interpretation according to which I am living as a Mayan Space Milkshake correct.) Nor can I decide, unless I am under a very specific delusion, to program myself as a mattress, since I know that that is not a way a person can be.

In light of this, we might think that there is a strong objectivity constraint on this sort of representation: the only things I can recognize as candidate programs for myself, and consequently the only things I can choose to program

myself with, correspond to socially constructed ways of life that are suitably recognized by the right sorts of people around me. But that constraint is too strong, since it would deny us all the possibility for innovation that goes along with being a serial hyperspecializer. (Not to mention, it would make it a mystery how we got such a large menu of programs to begin with.)

Still, the menu of programs that an agent has before him must provide *some* guidance. Recall Millgram's list of ways one can specialize: "VLSI engineer, comics inker, Cobra gunner, French professor specializing in eighteenth century poetry, adventure travel agent, director of cinematography". These are all quite specific programs, and some are very recent inventions. But notice that our names for them reflect the modification of more primitive types: a VLSI engineer is a particular kind of engineer, a comics inker is a particular kind of visual artist. This is the usual way that our programming innovation goes. We do not create new programs out of whole cloth. I'm not even sure how we would do this, since our understanding of the task of designing a solution to an ecological program is derived from our experience designing solutions to other engineering problems. Instead we look at the extant programs being run by our peers and modify them. Sometimes this involves abstract armchair engineering, more often it involves the kind of learning that Millgram calls practical induction. (Millgram 2016: 60-61) Importantly, there will also be limits to how dramatic this modification can be that are grounded in the interpersonal dimensions of self-programming. All the programs that Millgram lists depend on other people in obvious ways. The comics inker needs someone to write stories, the French professor needs a university, the Cobra gunner needs a pilot. Some of this dependence is material, like ink and helicopters. But the more important kind of dependence involves the recognition of other people. There is more to being a Cobra gunner than shooting Vulcan cannons; one must be *recognized* as having that competence by the right sorts of people. And there is more to being a French professor than professing about French; one must be *recognized* by students, colleagues, and university bureaucrats in the appropriate ways. As a matter of fact, nearly all the programs we might be interested in running will involve coordination with other people, which means that their individuation – what makes a person count as running one program rather than another – will depend on some manner of social construction. And this, in turn, will involve a coordinate recognition by those involved in the construction – by helicopter pilots, university bureaucrats, or one's community at large.

Even the programs that can be run in solitude are social roles in a more limited sense, and so depend on a kind of recognition. There is a subtle difference between the programs *Thoreauvian experimenter* and *anchorite*. To live according to the latter, I must do things, like praying or self-flagellating, that would

allow a hypothetical spectator to make the distinction. The difference between these modes of living do not reflect a brute fact – like the difference between gold and pyrite – but a collective understanding of the roles. So without this potential for recognition there would be nothing to make me count as running one program rather than another. But then in deciding to program myself as an anchorite, I must represent my chosen program as directing me to live in ways that *could* be so recognized by hypothetical spectators as anchoritic, at least in principle. In doing this I recognize those spectators as a constraint on what I can do.

The serial hyperspecializer's power of self-determination through self-programming depends on other people in two ways. Agents innovate by modifying a stock of programs created by other people, and the success of a modification depends on the recognition of others. This puts our programming endeavors under two competing pressures. There is a pressure toward innovation: we want more exact and less wasteful ways to fit more diverse niches that might better satisfy our unique and evolving needs and wants. But there is also a pressure toward conservatism and stability: too much innovation and no one is going to succeed in running their new-fangled programs because the coordination required would be impossible.

The same pressures can be found in the predicament of the artist. In art there is an obvious pressure toward innovation. The trite and familiar is not art, or at least not good art. But not anything goes. There are limits to how far one can carry one's novelty and invention and still be doing art. Art is to be experienced, and if no possible audience is apt to recognize some object or event as a work of art because it is simply too outlandish, then the work is not art.⁸ Thus it is inevitable that art will bear some imprint of influence from other people, be they predecessors or anticipated critics. Such a demand creates a countervailing conservative pressure. A famous example of this dualism of pressures is Harold Bloom's (1973) thesis about the "anxiety of influence": poets simultaneously strive for originality while recognizing the need for (and inevitability of) situating themselves within a poetic tradition, and this creates a distinctive anxiety about the signs of influence in their work that only the "strongest" poets can overcome.

These dual pressures toward innovation and recognition make art, as a kind, quite distinctive. It is unstable and subject to constant revision by the next genius and her reception. But this doesn't make it utterly formless – it

⁸ This idea gets codified into "institutional" analyses of art: a thing's status as art depends on institutions of art appreciation and criticism, i.e. on a sophisticated scheme of recognition. See, e.g., Danto (1964) and Dickie (1974).

doesn't mean that *anything* could be art. One way to characterize this peculiarity is to say that art is, to borrow W.B. Gallie's (1956) phrase, an *essentially contested* kind. (Gallie actually presents his thesis as concerning the *concept* of art, but this needn't worry us.) It is a kind whose nature is to be never be fully settled, but to instead be a focus of perpetual debate. Our concept *art* is bound up with persistent aesthetic debates, which, because of the kind of activity art is – “ever expanding, ever reviving and advancing values inherited from a long and complex tradition” (1956: 114) – we should expect to be open-ended. The concept, and so the kind it denotes, is therefore a forum for a debate that plays out through the advancement and consideration of rival conceptions of art – a debate that the practicing artist cannot help but participate in. And because this debate concerns an activity under the dual pressures I described, we shouldn't expect any definitive resolution. This is what makes art an essentially contested kind.

And so it goes with the kind agency – at least if Kant, Sartre, and Millgram are right. Or so I want to argue. Their accounts deny that certain constitutive features of agency are established facts, leaving that determination up to agents themselves. This opens up the possibility of radical innovation in how we live: the adoption of unprecedented programs, the valuing of things unmoored to instinct, the creation of new human essence. We might think that this potential for radical innovation makes human agency a completely boundaryless pseudo-kind – that *anything* could count as a manifestation of agency. But this can't be right. The agent's innovation depends on recognition by other people just as the artist's does. And that introduces limits on what I can do while qualifying as a human agent – as someone successfully programming themselves in the fashion of a serial hyperspecializer.

These boundaries are contested, though, just as the boundaries of art are. When we load a novel modification of a program, we are participating in the contestation of the kind agency. This process is analogous to the way the artist and her audience contest the kind art. In enacting a novel program, an agent exemplifies a conception of how one might adapt to a particular niche – a conception of agency – and tries to earn the recognition and cooperation (and, usually, approval) of others necessary to live out that conception. The agent needs the latter to establish her nascent program *as a program* in the way Cobra gunner, French professor, or comics inker are. In doing this she offers her program as something like an artistic exercise in living. Other people can resist, just as an artistic audience can resist, or they can play along. How they react will depend on their normative judgments about what is good and bad, sensible and silly – just as the artistic audience's judgments depend on their evaluative judgments – on how well our program fits its niche, and on our skill

at communicating the merits of the program. We should expect this process to be open-ended because there is no principled limit to the possible courses of innovation. And this makes agency an essentially contested kind.⁹

In saying that agency is an essentially contested kind, I am suggesting not that it lacks a constitutive nature, but that its nature is irremediably indeterminate. This indeterminacy is different from what we find with vague predicates, since it is not just about borderline cases. It is also different from any that might arise in the application of family resemblance concepts, since here the indeterminacy is a necessary consequence of central notion of agency rather than an incidental feature. Ultimately, this is what I think our third option from the trio listed above comes to: agency as a kind has a constitutive nature, but that nature is indeterminate because essentially contested.

I hope this thesis appeals to those who are skeptical about the ambitions of constitutivists. It strikes me as more plausible than the claim that *agency* is a family resemblance concept – that it is simply too heterogeneous to have a constitutive nature. But it does nonetheless seem capable of blocking constitutive arguments. If the nature of agency is indeterminate because essentially contested, then we won't be able to say of any particular principles or aims that one *must* adhere to them in order to be an agent. This is true not because we can produce some uncontroversial counterexample of agency being practiced in contravention of such aims and principles and therewith definitively establish that these principles and aims are *not* constitutive of agency. Rather it is true because controversy is inevitable, because any interesting principles or aims must be open to contestation by enterprising agents who might want to resist their authority by crafting innovative designs for living. We can resist these innovations, of course, as we can resist putative novelties in art. We just cannot do so preemptively and definitively by insisting that they *just are* contrary to the nature of agency. So constitutivist arguments in support of the normative authority of a particular principle will fail not because the principle fails to be constitutive of agency, but because it is necessarily an ongoing debate whether it is.

That said, this thesis may very well let in one of constitutivism's ambitions through the backdoor. Many constitutivists would like to show that morality has authority by dint of its connection to agency. If the thesis that agency is an essentially contested kind is correct, then we won't be able to do this for any particularly contentful moral *principle*, but we may be able to show that agents are committed to a pro-moral *activity*. For while the constitutive nature

⁹ This analogy between art and life becomes rather closer if we adopt a view on which the self is constructed through a quasi-artistic process such as narrative. For an overview of such views see Schechtman (2011).

of agency may not oblige us to adhere to any particular principles or adopt any particular aims, it does seem to require our participation in a particular activity, the activity of contesting the kind. Assuming I am right about the role of other people in our establishing our innovative modifications as bona fide programs, it would seem we can no more opt out of the contestation of agency than the artist or critic can opt out of the contestation of that kind. The artist who says that they are not merely contemptuous of the public, care only about posterity, or aim to scandalize the critics, but that they are *utterly* indifferent to the reception of other people *tout court* are not, I think, doing art. Whether they like it or not, the artist is participating in an old and ongoing debate about what art is.

So it goes with agency too. Someone who completely disengages themselves from the contestation of agency – who says they are not interested in agency – but are just *doing their own thing*, may not be able to program themselves in the way distinctive of serial hyperspecializers. If this is true, then whether they like it or not, agents are bound to participate in an old and ongoing debate about what agency is. What is important about this activity is that it requires the *recognition* of other people as having standing in how I behave. This kind of recognition is the essential ingredient in contractualist moral theories (e.g. Darwall 2006), so we might hope that a commitment to participating in the contestation of morality is ultimately a commitment to live according to programs that satisfy the demands of a more or less contractualist morality. Since I have left the details of *how* the recognition of other people is required for the agency of serial hyperspecializers quite obscure, I cannot offer any more than this sketch.¹⁰ But if it works out, then we can imagine a certain sort of quasi-constitutivist agreeing with those who are skeptical that agency has a fixed and immutable nature while still thinking that morality is ultimately grounded in the demands of agency. Such a person would hold that the authority of morality is grounded not in the nature of agency *as it necessarily is*, but in what we must do given what it isn't.

4. *Logic and taste*

So far I have focused on the implications of serial hyperspecialization for particular philosophical problems about agency, practical reason, and, in particular, the connection that constitutivists see between them. But this is only a small part of the significance that Millgram sees for the thesis. The greater issue, the one at

¹⁰ Though see Walden 2012 for an argument along these lines.

the heart of *The Great Endarkenment*, is the way that serial hyperspecialization threatens the way of life passed down to us from the Enlightenment.

Following Kant's famous formulation, Millgram says that the Enlightenment as we have received it is primarily a matter of autonomy: the ability to act and think for ourselves rather than being ruled by others or taking matters on simple faith. To be autonomous in this sense we needn't be self-sufficient autodidacts. We can rely on experts and the testimony of witnesses, but we should think it possible, at least in principle, to verify the soundness of their methods for ourselves. But autonomy, even in this idealized sense, is difficult in the age of serial hyperspecialization, Millgram says, because as hyperspecialization progresses, the differences between specializations become deeper. Specialists "develop proprietary systems of representation" and "internalize standards and guidelines that govern both their thinking about matters of fact and their choices of what to do", which are ultimately "unintelligible to non-specialists" (2016: 25-26). Taken to an extreme, this makes specialists in different specializations "logical aliens" (2016: 32-34). In a society where individuals are as specialized as all this, autonomy looks impossible. (It is also, obviously, a threat to the joint contestation of agency).

Millgram suggests that we might expect philosophy, the discipline traditionally concerned with Big Pictures, to provide the resources to bridge the gaps between alien specializations. Unfortunately philosophy itself has become so specialized that we cannot expect much from it. To do this work, it would require significant revision:

I recommend taking as our shared point of reference a repurposed and refocused philosophy of logic. Once we acknowledge that, from the perspective of anyone in our society, many of the other members of the society on whom we have to rely are logical aliens, philosophy of logic can be directed to a handful of related problems. [...] We need ways of assessing various modes of argumentation that have taken root in different disciplinary specializations [...] of assessing arguments constructed in an alien logic [...] of developing principled techniques for successfully identifying appropriate expertise [...] of managing the interfaces between disciplines [...] of managing arguments that traverse disciplines.

If we have a sufficiently ecumenical idea of the philosophy of logic, then this must be right. But Millgram goes on to offer a more specific and controversial suggestion for how this kind of work ought to be done. "The practice of this sort of philosophy of logic," he says, "ought to emphasize cognitive function analyses of available logical devices, and the design of improved replacements for them". (2016: 279) A cognitive function analysis, he explains, is a "design characterization of an intellectual device, along with an argument meant to

exhibit the work such a device does within a larger cognitive system and intellectual environment". (2016: 126) Millgram thinks that philosophy could bring about the necessary rapprochement between logical aliens occupying different specializations if it were concerned with the *work* that an intellectual device (a mode of inference or a standard of evaluation) does within a larger system and subsequently relied on this pragmatic information in answering the questions of the philosophy of logic.

This suggestion privileges what Millgram calls the design stance. In doing so it makes foundational a particular class of proprietary modes of representation, reasoning, and evaluation – the “logic”, loosely speaking, of engineering problems. I have two concerns with this proposal. First, I don’t see why the design stance deserves this exalted status. Surely there are other stances that could do similar work – other distinctive modes of evaluation and reasoning through which we can size up the panoply of specializations around us. Why the design stance instead of one of these? My second concern is that the suggestion is too hegemonic. In requiring that alien specializations pass muster as solutions to design problems, we are not reckoning with them on their own terms but according to the distinctive standards of the design stance. While we are not quite demanding a reduction of alien logics to the design stance, we are demanding an interpretation of those logics that meets criteria native to the design stance. This threatens to beg the question. For there will be some specializations fitted to particular niches for which the proprietary logic of design problems and solutions is itself flagrantly inappropriate. Think, for example, of the specializations of the aesthetes in *Patience* or the bohemians in *La Bohème*. To criticize the “intellectual devices” of these self-consciously unpragmatic specializations on the grounds that they are not useful would seem to beg the question against them. So without an argument for why the design stance ought to be privileged over others, using it in this way seems inappropriate.

This puts us back in the lurch. If not through cognitive function analysis, how can the necessary philosophy of logic proceed? A moment ago I suggested that serial hyperspecialization placed agents in the same predicament as artists. I now want to extend this analogy one step further by suggesting that the philosophy of logic Millgram envisions ought to begin as an exercise in *taste*.

Kant observed that judgments of taste are characterized by an unusual pair of properties. The first is normativity. When I judge that a poem is beautiful, I am not merely reporting my internal pleasure, but taking myself to have gotten something right. This puts me into conflict with someone who holds a contrary judgment in a way not shared by judgments of “mere agreeableness”: we have a genuine disagreement, and one of us is in error. The

second property is taste's unruliness. My judgment that the poem is beautiful is not simply the application of a rule, as my judgment that something is a table or a square is. If taste worked like this, then the construction and appreciation of beautiful poems would be the essentially mechanical enterprise of applying the correct rules of taste. But this is manifestly not how artistic production and criticism go, nor do we have any reason to believe that there are any rules that could successfully codify beauty. So taste cannot depend on rules. (1790, Eng. tr. 2000: §§6-8)

This pair of properties perform an important task for aesthetic kinds. It allows such kinds to be contested while remaining unified as a single kind. If aesthetic judgments did not make a claim to universal agreement – if they were not normative – aesthetic disagreements would naturally lead to fragmentation of the kind. If you have a conception of art on which *Götterdämmerung* is the paradigm, and I have a conception on which Greek pottery is, then we may be tempted to take ourselves to be talking about two completely different kinds, Wagnart and Potterart. The fact that we also disagree about whether *Götterdämmerung* is flaccid or transcendent and whether Greek pottery is lifeless or graceful brings us back into conversation and forces our conceptions of art to be genuine rivals for a common kind. On the other hand, if these aesthetic judgments were based on rules, then we might reasonably suppose that those rules could deliver definitive answers to questions about the boundaries of art and thereby render the kind uncontested. But the absence of such rules at least makes it an open possibility that the activity of contesting the kind is truly open-ended. So art's status as a singular but contested kind is made possible by these unusual features of our evaluation of art.

How does this feature of taste help us with Millgram's problem? At bottom the task that Millgram sets for the philosophy of logic is the evaluation of what he calls alien logics. We need these evaluations if we are to avoid being forced into heteronomy by the fragmentation of specializations. This evaluation seems so difficult because most evaluations presuppose standards and rules which are themselves linked to the proprietary systems of representation, inference, and assessment. So if we venture an evaluation of alien logic A, it will likely be premised on the standards of some proprietary system B, which, in a state of hyperspecialization is likely to be alien to A. This exposes us to the kind of objections I lodged against Millgram: that we are illicitly privileging system B and begging the question against A. One way around this difficulty would be to employ a mode of evaluation that was not premised on a rule, and so not tied to a particular proprietary logic. And this is exactly what taste offers us: evaluations that are genuinely normative but not rule-governed. Of

course, taste judgments can be just as biased against the alien, but this bias is not *de jure* and can be overcome by good critics. Taste offers the possibility of a non-hegemonic evaluation of an alien logic.

These evaluations will seldom ascribe beauty or ugliness. Instead they will traffic in quasi-aesthetic thick concepts. “Specialization A’s standards of assessment are shallow and unsophisticated”. “The intellectual devices employed by specialization B are sloppy”. Offering such verdicts is only the beginning of our evaluation, however. We have to defend and explain them, in much the way the critic defends and explains her evaluations: by presenting features of the target – in this case the machinery of alien specializations – in a way designed to elicit the same spontaneous judgment in our audience. Sometimes we can do this by trying to show that these devices are useless or inefficient, sometimes by showing they are tedious. It is the critical conversations that ensue from these verdicts that I propose offer the possibility of a forum for logical aliens to offer evaluations on neutral ground.¹¹

Consider an example I mentioned a moment ago. Suppose that Victorian aestheticism of the sort we see in *Patience* amounts to a genuine specialization with its own distinctive “logic”. I said that if we criticize this specialization through a cognitive function analysis, we could be accused of begging the question. It seems very odd to decide on the merits of aestheticism by asking what problems it solves. But Gilbert and Sullivan’s satire of the movement cannot be accused of this. It does not apply an alien standard to the aesthetes. It only presents a stylized version of their way of life designed to elicit a judgment of taste: that aesthetes are *very silly people*. This verdict is not the end of our evaluation; it is only the beginning of a conversation between logical aliens (including, perhaps, aesthetes and engineers) over the merits of this way of life – a conversation that does not presuppose either’s native standards.

The aim of this critical practice is not to reach the right answers about which specializations are worth countenancing. All we can hope for is fostering a discursive community in which people can talk across the logical chasms of specialization. Doing this should suffice to stave off the especially radical kind of fragmentation that hyperspecialization threatens us with. What I have suggested here is that judgments of taste have the formal characteristics that make them suited to this task. If I am right, then the savior of the Enlightenment will not be the cognitive engineer who can translate alien logics into the framework of design problems, but the critic who can evaluate without translation. They will not be someone with the right box of standards and rubrics, but of a person “strong sense, united to delicate

¹¹ On this point see Isenberg (1973) and Nehamas (2007: 97ff).

sentiment, improved by practice, perfected by comparison, and cleared of all prejudice” (Hume 1757: 229). For these are the skills we need to make fair, persuasive, and productive normative evaluations of alien ways of life (and novel works of art) without begging the question.

Kenneth Walden
Dartmouth College
kenneth.e.walden@dartmouth.edu

References

- Bloom, Harold, 1973, *The Anxiety of Influence: A Theory of Poetry*, Oxford University Press, New York.
- Danto, Arthur, 1964, “The artworld”, in *Journal of Philosophy*, 61, 19: 571-584.
- Dickie, George, 1974, *Art and the Aesthetic: An Institutional Analysis*, Cornell University Press, Ithaca.
- Darwall, Stephen, 2006, *The Second-Person Standpoint*, Harvard University Press, Cambridge.
- Enoch, David, 2006, “Agency, shmagency”, in *Philosophical Review*, 115, 2: 169-198.
- Enoch, David, 2011, “Shmagency revisited”, in M. Brady, ed., *New Waves in Metaethics*, Palgrave MacMillan, New York.
- Hume, David, 1757, “Of the standard of taste”, in *Four Dissertations*, Millar, London.
- Ferrero, Luca, 2009, “Constitutivism and the inescapability of agency”, in R. Shafer-Landau, ed., *Oxford Studies in Metaethics*, vol. 4, Oxford University Press, New York.
- Foot, Phillipa, 2003, *Natural Goodness*, Oxford University Press, New York.
- Gallie, W.B., 1956, “Art as an essentially contested concept”, in *Philosophical Quarterly*, 6, 23: 97-114.
- Isenberg, Arnold, 1973, *Aesthetics and the Theory of Criticism*, Cornell University Press, Ithaca.
- Kant, Immanuel, 1781/1787, *Kritik der Reinen Vernunft (Critique of Pure Reason)*, Verlag Hartknoch, Riga; Eng. tr. Paul Guyer and Allen Wood 1998, Cambridge University Press, New York.
- Kant, Immanuel, 1786, „Mutmaßlicher Anfang der Menschengeschichte“ (“Conjectural beginnings of human history”), in *Berlinische Monatschrift*, Berlin, Eng. tr. Pauline Kleingeld 2006, *Toward Perpetual Peace and Other Essays*, Yale University Press, New Haven.
- Kant, Immanuel, 1790, *Kritik der Urteilstkraft (Critique of Judgment)*, Ligarde und Friedrich, Berlin, Eng. tr. Paul Guyer et al. 2000, Cambridge University Press, New York.

- Katsafanas, Paul, 2013, *Agency and the Foundations of Ethics*, Oxford University Press, New York.
- Korsgaard, Christine M., 2003, *Self-Constitution: Agency, Identity, and Integrity*, Oxford University Press, New York.
- Millgram, Elijah, 2010, "Pluralism about action", in T. O'Connor *et al.*, eds., *A Companion to the Philosophy of Action*, Wiley-Blackwell, Malden.
- Millgram, Elijah, 2015, *The Great Endarkenment*, Oxford University Press, New York.
- Millgram, Elijah, 2016, "Practical reason and the structure of actions", in E. Zalta, ed., *Stanford Encyclopedia of Philosophy*, Summer 2016 edition, URL = <https://plato.stanford.edu/archives/sum2016/entries/practical-reason-action/>
- Nehamas, Alexander, 2007, *Only a Promise of Happiness*, Princeton University Press, Princeton.
- Schechtman, Marya, 2011, "The narrative self", in S. Gallagher, ed., *The Oxford Handbook of the Self*, Oxford University Press, New York.
- Silverstein, Matthew, 2015, "The shmagency question", in *Philosophical Studies*, 172, 5: 1127-1142.
- Sartre, Jean-Paul, 1943, *L'Être et le néant (Being and Nothingness)* Gallimard, Paris; Eng. tr. Hazel Barnes 1956, Philosophical Library, New York.
- Tiffany, Evan, 2012, "Why be an agent?", in *Australasian Journal of Philosophy*, 90, 2: 223-233.
- Velleman, J. David, 2009, *How We Get Along*, Cambridge University Press, New York.
- Walden, Kenneth, 2012, "Laws of nature, laws of freedom, and the social construction of normativity", in R. Shafer-Landau, ed., *Oxford Studies in Metaethics*, vol. 7, Oxford University Press, New York.
- Wood, Allen W., 1999, *Kant's Ethical Thought*, Cambridge University Press, New York.