Expertise and the Fragmentation of Intellectual Autonomy

C. Thi Nguyen

Abstract

In *The Great Endarkenment*, Elijah Millgram argues that the hyper-specialization of expert domains has led to an intellectual crisis. Each field of human knowledge has its own specialized jargon, knowledge, and form of reasoning, and each is mutually incomprehensible to the next. Furthermore, says Millgram, modern scientific practical arguments are draped across many fields. Thus, there is no person in a position to assess the success of such a practical argument for themselves. This arrangement virtually guarantees that mistakes will accrue whenever we engage in cross-field practical reasoning. Furthermore, Millgram argues, hyper-specialization makes intellectual autonomy extremely difficult. Our only hope is to provide better translations between the fields, in order to achieve intellectual transparency.

I argue against Millgram's pessimistic conclusion about intellectual autonomy, and against his suggested solution of translation. Instead, I take his analysis to reveal that there are actually several very distinct forms intellectual autonomy that are significantly in tension. One familiar kind is *direct autonomy*, where we seek to understand arguments and reasons for ourselves. Another kind is *delegational autonomy*, where we seek to find others to invest with our intellectual trust when we cannot understand. A third is *management autonomy*, where we seek to encapsulate fields, in order to manage their overall structure and connectivity. Intellectual transparency will help us achieve direct autonomy, but many intellectual circumstances require that we exercise delegational and management autonomy. However, these latter forms of autonomy require us to give up on transparency.

Keywords

Expertise; autonomy; social epistemology; trust; testimony; epistemic dependence

In *The Great Endarkenment*, Elijah Millgram offers a vision of a very particular sort of intellectual apocalypse. This apocalypse is due to the onset of intellectual hyper-specialization. Human knowledge has splintered across a great many distinct fields of inquiry. These fields are mutually incomprehensible to one another — each field requires not only its own technical jargon and specialized knowledge, but its own specialized form of reasoning. What constitutes a good argument varies radically from field to field. Thus, even a well-educated expert in one field will not be able to distinguish real experts from fake ones in another field. And, if it isn't already obvious, Millgram thinks that this apocalypse has already arrived.

Hyper-specialization presents an immediate problem for anybody who hopes to achieve complete intellectual autonomy — that is, who hopes to understand and control all the elements of their beliefs for themselves. Moreover, the crisis of hyper-specialization threatens successful practical activity itself. This is, says Millgram, because practical arguments are draped across a large number of hyper-specialized areas. A single argument for a particular conclusion ("use this kind of shielding in your nuclear reactors") has premises in one field (particle physics), inferences from those premises performed in another field (chemistry), and then conclusions in a third field (nuclear engineering). Thus, no single person has control of all the steps of the argument. This draping of arguments across fields gives us a three-pronged version of the problem of intellectual autonomy. First, says Millgram, no individual can understand all the premises of such a draped argument for herself. Second, even if we are permitted to deputize other people in our quest for knowledge, we do so in the hopes that they will perform as we would have. But experts in different fields go about their intellectual business in such radically different ways that autonomous deputization seems impossible. Third, says Millgram, every argument has defeaters, and as

we extend an argument over many fields, we lose the ability to recognize when an argument has met a defeater. What we need, says Millgram, is to manage the interface between different fields, and one of the groups best able to do this would be analytic philosophers trained in conceptual analysis. Such conceptual analysts could act as the translators between different fields, easing the problems engendered by hyper-specialization (Millgram 2015: 27-44).

I'm going to assume that Millgram is right about the basic infrastructure of his intellectual apocalypse: that there are distinct fields that are mutually incomprehensible to each other. This means that an expert in one field cannot directly assess the quality of an argument from another field. Are there other means by which we can still come to trust other experts and other fields, even if they are incomprehensible to us? I will consider Millgram's three problems separately, and will offer some solutions that might, when considered independently, ease our local worries. But these solutions will turn out to run at cross-purposes with one another. Millgram's account, I will argue, does not show that intellectual autonomy is impossible. Rather, it will force us to realize that there are different kinds of intellectual autonomy, and that they are often in tension with one another. To support one kind of intellectual autonomy, we should seek greater intellectual transparency between fields. But other interests will call for different sorts of intellectual autonomy, which, in turn, will call for the opposite of transparency: an aggressive compartmentalization between fields.

What we learn is that pressure from the expanding scope and complexity of human knowledge has fragmented intellectual autonomy — or rather, it reveals that intellectual autonomy has always been fragmented. In the past, the relatively small size of human knowledge let us hold onto the hope of achieving of a unified form of intellectual autonomy. The rapid growth of the sciences forces us to admit that there are different forms of

intellectual autonomy, that they are in tension with each other, and that we must sometimes choose between them.

The problem of expert recognition

How might we hang on to some kind of intellectual autonomy in the era of hyper-specialization? How much can we really think for ourselves anymore? Once upon a time, we might reasonably have hoped to achieve *direct epistemic autonomy* — to understand all the reasons and evidence that support our beliefs entirely for ourselves. But the size of human knowledge, and the mutual incomprehensibility of fields, seems to make direct epistemic autonomy impossible. Complete intellectual independence is a pipe dream in any world with rich domains of expert knowledge (Hardwig 1985). But perhaps we can achieve a weaker kind of autonomy — call it *indirect epistemic autonomy*, or meta-autonomy for short. Meta-autonomy is a weaker requirement; to achieve it, an agent need not gather all the evidence and reasons directly for herself. She need only autonomously gather evidence and reasons about whom she ought to trust, and then use that autonomously generated trust to filter information from the various purported experts.

Millgram allows that such a distanced and indirect sort of meta-autonomy might be worth having. Says Millgram: even if we have given up on the hope of absolute epistemic autonomy, perhaps we could at least take responsibility for the choice of experts (Millgram 2015: 29-30). But, says Millgram, even meta-autonomy is impossible to achieve, since you need to already be an expert in order to successfully locate other experts (30, 44-8). This sort of argument may be familiar from recent work in the epistemology of testimony, especially in moral epistemology, where it has been called the *credentialing problem* (though the problem is as old as Socrates) (Cholbi 2007; LaBarge 1997; McGrath 2011). Here's the worry: an

expert is marked by their capacity to produce good arguments in their field; but if we cannot tell good arguments from bad in some field, then we cannot recognize the experts in that field. So, in precisely those cases in which we truly need experts, we will be unable to find them. And surely, in the era of hyper-specialization, we cannot sort good arguments from bad in distant fields.

I will argue, contra Millgram, that we can, in fact, achieve certain forms of meta-autonomy, but that achieving it involves sacrificing some of the goods that are traditionally associated with direct intellectual autonomy. Let me sketch what a reasonable meta-autonomy might look like. Suppose that I have no expertise in a given field. How might I go about finding the experts? Experts are not simply the people that produce good arguments; they are the people that reliably produce good results. In some fields, good results are easy to evaluate, even without any expertise. Those fields possess what we might call a litmus test — a success-condition that's recognizable to the inexpert. In the most *obvious* fields, that litmus test is applicable on an individual and immediate basis. If you claim to be an expert axe-thrower, I can test your claim by handing you an axe and indicating an appropriate target. Of course, most fields are more *subtle* — the expert is she who produces a good result, but the proper assessment of good results itself requires expertise. And pessimists about the possibility of finding moral experts are usually quick to point out that, while many empirical fields have a clear litmus tests, moral expertise does not (Cholbi 2007, 325-32; McGrath 2011, 96-9). (For an extended discussion of litmus tests, and subtle and obvious fields, see (Nguyen 2018)).

Surely there are litmus tests for the more straightforward empirical domains. The domain of knowledge concerning skilled axe-throwing has a clear litmus test. But, just as surely, there are more obscure empirical domains that lack any straightforward litmus tests. Good abstract

expressionist painting, good proofs in number theory, and good work in particle physics are all incomprehensible to the layperson.

Of course, such fields produce their own internal rankings. One can, perhaps, look to see who the good economists are by seeing who publishes in the most prestigious economics journals, or has won the Nobel Prize in Economics. But this simply defers the problem of expert recognition from assessing individuals to assessing whole fields, for now we have to distinguish pseudo-fields from legitimate fields. Fields can become corrupted and epistemically wayward. Think, for example, of the experts that have been selected and officially licensed by Scientology. (I suspect that most academics would be willing, over a private drink, to finger certain other academic fields as corrupt.) But there are ways around this problem. First, some fields, which do not admit of individual litmus tests, may be amenable to field-wide litmus tests. I cannot understand the individual pronouncements of an aeronautical engineer, but I trust the field of aeronautical engineering in general because planes do not generally fall out of the sky. Other sorts of litmus tests are possible — whole fields have been discredited when top journals publish purposely fraudulent gibberish (Millgram 2015: 32). Some fields do not admit of individual litmus tests, but do admit of collective ones. And, to the extent that these fields have internal procedures for ranking, then trust in a whole field can lead to trust in that field's certification procedures, which, in turn, can lead to trust in individual experts. In this case, we have achieved, not direct autonomy, nor even meta-autonomy, but, at best, meta-meta autonomy. But at least that's something.

Furthermore, though some fields don't lead to a direct litmus test — theoretical physics, for example — we can still perform what Philip Kitcher calls "indirect calibration" (Kitcher 1993: 320-3). Fields overlap. If I trust the engineers, and engineering substantially overlaps with mathematics, then I can trust the engineers via a litmus test, and then trust the

mathematicians via the engineers. Nuclear engineering has a fairly vivid litmus test, and the field depends on the results of theoretical physics. Thus, my trust in nuclear engineering can be extended to trust in theoretical physics. It may be that some fields are not hooked up in this way — that they are cognitively isolated. But Millgram's primary concern is with fields that do, in fact, hook up with other fields to produce, eventually, practical outcomes — so the fields that Millgram worries about are precisely those in which indirect calibration to a litmus test is possible.

For a linked field, you don't need to be an expert to have *some reason* for believing another to be an expert. You can use a publicly available litmus test, and then trace a line of trust through some number of fields. Why does this matter? Well, it depends on what you wanted out of autonomy in the first place. If your goal was radical and total intellectual independence — well, then Millgram's surely right, and you're completely out of luck in the modern world. But suppose your goal were more modest: it is that you have reasons, of some sort, under your control for your beliefs, rather than having to adopt them out of unthinking obedience (Nguyen 2010). In that case, indirect calibration and litmus tests give us *some* reason. If the contrast class to intellectual autonomy is intellectual servility, then there are more and less intellectually autonomous ways of going about gathering testimonial knowledge in a hyper-specialized world. One would have no intellectual autonomy whatsoever if one simply accepted claims to intellectual authority at face value without any further reflection — this would be complete intellectual servility. In contrast, achieving some degree of meta-autonomy is surely better: one could conduct a reasoning process of one's own, in which one thinks about the litmus tests and calibrations and comes to an individual assessment of entire fields, and then trusts those fields' internal rankings. And such a process is not just rote. If one had meta-autonomy, one could actively manage one's trust in other

fields. One could come to reject fields for failed their litmus tests, even though they present themselves as, or are widely believed to be, good fields. Daphne Miller, for example, notes that even a layperson should have good evidence that nutrition science is not a trustworthy field, simply from the frequency with which it reverses itself and condemns its own recent declarations (Miller 2008).

Why might we want this form of indirect autonomy? Perhaps it is a good in and of itself. But one might also think it has practical epistemic value. Philip Nickel has suggested that one reason to want intellectual autonomy is adaptability to changing conditions. When one simply accepts a belief on authority, one has no grounds for adjusting that belief as conditions change. If I simply, for example, memorized from a textbook the belief that egg yolks were bad for you (as was generally believed a few decades ago), my belief wouldn't update as new evidence came up in the field. Similarly, as Nickel points out, if I go to a moral guru and get advice about a moral dilemma I'm in, and carry away a simple directive, my actions won't respond to any changes in the situation. Individual contact with the reasons for one's belief provides one with the capacity for adapting one's beliefs (Nickel 2001). Notice how this plays out with distanced indirect calibration. Suppose there is a field of, say, theoretical Jungian macroeconomics, which I had, at first, trusted. After all, there are whole academic departments that specialize in it, and even a specialty journal. As an outsider, I have no capacity to adjust my beliefs independently about individual pronouncements, or even about the reliability of particular, within Jungian macroeconomics. But if I eventually come to notice that all the fields that draw from it are starting to fail their litmus tests, then I can let my distrust filter up and come, via indirect calibration, to distrust Jungian macroeconomics. This is a very distanced sort of autonomous control, but it may be the best of which we are capable.

We have found the credentials problem is not entirely insurmountable. We have some limited resources to establish indirect intellectual autonomy — more, at least, than Millgram strictly permits. Still, the distanced sort of meta-meta-meta-etc-autonomy we are actually capable of achieving might seem deeply unsatisfying. Somebody who desired direct autonomy for its own sake have reason to eliminate the links and bring those distant fields under their own understanding whenever possible. Furthermore, somebody who desired intellectual autonomy for Nickel-style reasons would also want to decrease the number of links. The fewer insulating layers between myself and the argument, the more adaptable I will be, and the more control I will have over the premises and steps of the arguments I use. How might we decrease the indirectness of our autonomy? Millgram has suggested that one way to ease the autonomy problems of hyper-specialization is to set an army of philosophers at the interfaces between fields. They could help experts to understand each other's fields by providing greater and clearer translations between the fields. In short, a desire to decrease intellectual distance will lead to a desire for greater intellectual *transparency*. Our army of philosophers will, by providing conceptual translations and increasing intellectual transparency, serve to increase the intellectual autonomy of experts in one field who needed to rely on the work of experts in another field. So long as direct autonomy is the ideal, and indirect autonomy a mere compromise, we should seek more to minimize trust and maximize transparency.

Intellectual autonomy and proxies

Let's turn to the next of Millgram's problems for intellectual autonomy — that other people reason differently from ourselves and that this creates a particularly great problem

for managing defeaters across fields. Though he broaches these topics together, I think they are worth thinking about separately. What I hope to show, in these next two sections, is that some interests we have which bring us to want intellectual autonomy can only by satisfied by sacrificing transparency and translation.

Let's start with the idea that different specialists reason in radically different ways. Millgram suggests that people in other fields are "logical aliens" – people whose notion of what counts as a good argument is different from your own.

> However, when you delegate part of your deliberation to a logical alien - put less dramatically, to a specialist whose idea of what a good argument is differs from yours - what he comes up with will not normally conform to standards you accept. When an outsider is aware of another discipline's internal standards, he may well and is even likely to think they are wrong headed... Because you're not delegating to someone who thinks as you do, we no longer have an explanation for how delegation of this sort preserves your autonomy — or, perhaps more carefully, we haven't yet elaborated a notion of autonomy that makes room for such an explanation. (Millgram 2015: 33-4)

Implicit in this passage is an account of how intellectual delegation could be autonomy-preserving. If I have good reason to think that another person would reason just as I do, then offloading a reasoning task to them would be autonomy-preserving. This condition is stronger than what has emerged previously. In the earlier discussion of meta-autonomy, we had sought only to discover through our own understanding that some other expert was reliable; we had not sought to establish that they also thought as we did. (The reason I trust my accountant is precisely because he does not think as I do, for I am lousy at accounting.) But this new condition adds the requirement that, when we delegate our deliberation to

another, that this other person would must *reason in the way that we would have* in order for that delegation to be autonomy-preserving. Let me call this the *proxy requirement.*[1]

In the science fiction novel *Stations of the Tide*, Michael Swanwick imagines a future in which a bureaucrat, when pressed for time, can create mental copies of himself, and set each to its own task (Swanwick 2011). This would, I take it, result in the most perfect possible intellectual proxy. It's absolutely certain that the proxy would do things as I do, and it seems intuitively plausible that the use of such a perfect proxy would preserve my autonomy. In the novel, the bureaucrat absorbs the memory of each proxy once it has finished its task. Let me modify the scenario a little for my purposes. Imagine that I can make mental copies which are not reabsorbed whole into my mind, but simply present me with their conclusions, shorn of the evidence and reasoning. There is, after all, a limit to how much my brain can hold. Even then, if my perfect proxy simply reports to me a conclusion, stripped of their reasons — still, it seems like intellectual autonomy is preserved. After all, this is the relationship we have with past versions of ourselves. I have, for example, decided for myself that there is something deeply wrong with Hegelianism, though I performed the reasoning over a decade ago and cannot remember any of the details. But trust in my past self still seems autonomy-preserving. I have good reason to think that old-me would have reasoned quite similarly to present-me. And, if I wanted, I could re-open my old Hegel books and think it through for myself again. But the fact that I have not, and simply rely on my memory of my past self's eventual conclusion, without any memory of my past self's reasoning, doesn't undermine my intellectual autonomy.

---

[1] Note that Millgram himself, in the quoted passage, floats the proxy requirement and then distances himself from it, allowing, in the end, for the possibility of a reformulated notion of autonomy that might drop the proxy requirement.

Leaving the science fiction, it seems plausible to think that imperfect proxies are autonomy-preserving to the degree that they resemble perfect proxies. The more I can think that somebody would do something precisely as I would do it, the closer I can get to that dream of complete, entirely direct intellectual autonomy. But, an important question: do I really want perfect proxies if I am simply in the business of getting things right? That doesn't necessarily seem to be the case. So long as we think that there isn't one best way to reason — that different ways of thinking might be better adapted to different terrains — then, for purely epistemic reasons, I might not want to be confined to using only perfect intellectual proxies. I would not, for example, want somebody who thought like a philosopher to give me marriage counseling advice, oversee a non-profit charity, or to guide me through free jazz. In many cases, I don't seek anything like an intellectual proxy at all. I am, instead, seeking somebody that I can trust who is quite different from me — somebody, in fact, that I need to trust precisely *because* they are quite different from me. The vast difficulty of intellectual life in the hyper-specialized world isn't simply that I lack for time; it is that different hyper-specializations seem to require vastly different intellectual makeups. If you don't believe this, then I invite you to imagine a world in which music, television, couples therapy, military leadership, restaurant cooking, and children's education were all in the charge of analytic philosophers.

What could the basis for such trust be, then, if not proxy-hood? It may vary from terrain to terrain. In some cases, as with a marriage counselor, I trust that they have my best interests at heart. With my guide to the world of free jazz, I trust that they have good taste and are sensitive to the needs of a newcomer. But in most academic and scientific terrains, I am trusting in their ability to get it right. This may include trusting in their intellectual integrity, sensitivity to facts, intellectual verve, or whatever characteristics happen to be

useful for their field. Notice that if this is true, then the difficulties of hyper-specialization pull us in two very different directions. In order to maximize our intellectual autonomy, we would want to seek perfect proxies, or their closest available approximation. But because the demands of different fields are not simply on our time, but demands for different cognition, then our drive for the greatest accuracy leads us to seek something very different.

To return to Swanwick's tale: suppose, in the same science fictional universe, that I was aware of the vast differences between fields at a young age. So, cleverly, before I intellectually specialized, I made many copies of myself, enough for every intellectual discipline, and sent them each to graduate school in a different field, to each be changed into the kind of logical alien that their respective field requires. But if my copies were trained into different fields, then they would no longer really be intellectual proxies for me — not in Millgram's sense, at the very least. They certainly will not go about the task exactly as I would have. But what, then, was the point of the exercise? Perhaps that I know, at least, that I can trust them — I believe they are well-intentioned in their reasoning, if not perfect proxies for how I would have reasoned.

Annette Baier suggests that trust is essentially making yourself vulnerable to somebody else — it is entrusting something of yours to their care (Baier 1986). When I trust a doctor, I make my body vulnerable; when I trust a romantic partner, I make my emotions vulnerable. When I trust another academic and rely on them, I make my belief system vulnerable. In the case of perfect proxies, the trust that is required of me is rather thin. I am deputizing those who think as I do — if they fail, they failed because they haven't reasoned well in this particular case, but at least the kind of reasoning they will engage in has been chosen by me. I am only vulnerable to a simple failure of applied reasoning. But when I trust a logical alien, my trust is significantly thicker. I am trusting them not only to reason well, but to have

13

adopted the form of reasoning best suited to the task, even if that form of reasoning might not seem good to me. This is a vaster and more uncomfortable form of intellectual vulnerability.

But if we do think that different fields require different reasoning — that is, if we think that logical aliens have arisen not by accident, but because cognitive life demands significant intellectual diversity — then this more profound trust is actually requisite for getting things right in a world of hyper-specialization. And, insofar as intellectual autonomy consists, at least in part, in translating my interests into suitable action through my own activity, then the capacity to delegate in this manner is helps me instantiate a form of intellectual autonomy. Moreover, it is a form of autonomy which runs counter to the interest in approximating direct intellectual autonomy through perfect proxies. A kind of meta-autonomy is still available here — I may choose to trust people and fields for good reasons. I might, for instance, choose to trust a logical alien because they seemed to have some sort of integrity, or a real love of the truth. But it is a kind of indirect autonomy that is quite at odds with any hope of transparency, for the very reason I wish to trust them is that their capacities and abilities are substantially different from my own.

This, let me suggest, is significantly different kind of intellectual autonomy. It is another way in which we can seek a relationship to knowledge and others, over which we might a distinctive form of control. It is the capacity to invest others with intellectual trust, even when their intellectual work is not transparent to us. Robert Nozick has suggested that something very interesting happens in love relationships: we begin to delegate intellectual tasks to one another. I read all the political news, and my wife reads all the environmental news, and we report and digest any extremely important information for the other. And if the other person doesn't report that anything interesting has happened, I trust them (Nozick 1990). Again, here is a kind of intellectual delegation and trust, but it doesn't involve anything like a perfect

proxy. In fact, the reason that I read the political news and my wife reads the environmental news is that I am a philosopher and she is a scientist. Other relationships like this can occur elsewhere. I recently co-authored an article with an expert in the field of game design. I knew the philosophical literature about the nature of games and play, and my co-author was buried in the literature of computer game design. We wrote the paper together precisely because neither of us had the time to read that other field entirely. The paper emerged out of a long set of conversations between us, and we both signed our names to it. Now the details of what happened there are very complicated, but we could describe what's going on as accessing a distinctive kind of intellectual autonomy, very different from any of the familiar old individualistic and direct forms. In this kind, I am autonomous because I have chosen and delegated responsibility to somebody out of an active sense of trust and because I treat myself as responsible for the whole. Let's call this *delegational autonomy*.

Direct autonomy leads us to want transparency. When we turn to meta-autonomy, Millgram suggests that we ought to increase transparency as much as possible, to approximate direct autonomy. But I've argued that, in some cases of meta-autonomy, we ought not wish for transparency. We ought, instead, wish to have delegational autonomy. We have delegational autonomy in those intellectual processes which are under our control, and regulated by our reasons, but during which we place our trust in people whose capacities are beyond what we can understand for ourselves.

The problem of cross-disciplinary defeater management

Let's turn now to the third of Millgram's problems arising from hyper-specialization. Even if we manage to locate the right experts, and even if we are empowered to trust them,

another problem arises: what Millgram calls the problem of cross-defeater management.

Every successful argument, says Millgram, has its defeaters — conditions which would show

the argument does not, in fact, go through.

> Doing a decent job of thinking for yourself requires sensitivity to the defeating
> conditions of the argument supporting your views, and if you are not an expert in the
> subject matter of those arguments, you do not generally control those defeating
> conditions… When you are assembling a defeasible argument, you should be
> confident in your conclusions only to the extent that you control the argument's
> defeasibility conditions; in a less fancy way of saying it, your confidence should not
> outrun your ability to catch problems and bugs in your argument. (Millgram 2015, 31-
> 2)

The problem comes when we take an argument from a field outside our expertise and

then export it elsewhere and apply it on our own recognizance. An expert in one field cannot

be expected to control all the defeaters of an argument from another field; we should expect,

when we export arguments, to miss applicable defeaters. But modern practical arguments are

draped across different fields. In order to apply them, we must borrow chunks of arguments

from other fields and apply them in our own. But since we know that we will miss any

defeaters to those exported chunks of argument, we should have very low confidence in the

success of such cross-disciplinary arguments. This Millgram dubs the problem of cross-

disciplinary defeater management.

Perhaps an example will help. A physical therapist might recommend to me the exercise

of shoulder dislocates for stiff shoulders, especially since I cannot raise my hands directly

above my heads (Despite the threatening name, this is simply a simple overhead rotational

movement with a broomstick.) I might find that this exercise helps, and start teaching my

friends the exercise. However, I am not aware of the following defeaters: the exercise is only

useful for shoulders that are stiff from being hunched over a laptop, and not useful for

shoulders that are stiff from the impact of a car accident. Furthermore, the exercise is useful

for most people, but in the rare case of a particular kind of nerve impingement, will make it worse. My lack of sensitivity to defeaters will lead to misapplications, but my lack of sensitivity is understandable, since I am not a physical therapist. Versions of this problem apply, Millgram suggests, whenever an expert in one hyper-specialized domain has to apply results from another hyper-specialized domain.

I take the problem of cross-disciplinary defeater management to be the most threatening of Millgram's problems. Notice that none of the potential solutions to the other problems I described above will work here. It does not matter even if we manage to identify true experts in other fields with absolute precision. Even if I have found the true expert and received their correct argument, the mere fact that it is I, and not the expert, that has to apply that argument to new situations, leads to the cross-defeater problem (32).

Mere translational work between fields won't relieve this problem. It doesn't matter if I understand the concepts from another field; without actual expertise in that field, I won't control the defeaters from arguments from those other fields. In fact, greater transparency may simply exacerbate the problem. Transparency offers the illusion of control. It offers access to seemingly good arguments from other fields, while hiding the complexity of possible defeaters. It is an invitation to failure. Good delegation won't work; even if I have perfect knowledge of my field, and you have perfect knowledge of your field, and we have chosen to trust each other for very good reasons, the key information is lost when our expertise is split between two different people. Millgram suggests another solution: that individuals be trained in multiple fields (280). If I, for example, have a dual PhD in physics and in chemistry, then I may be able to manage defeaters across both these arguments. But notice this only solves the problem for those special individuals — it does not solve the problem for any of the rest of us, who only have been trained in one field. We may consult one

of these special doubly-trained individuals to look over our arguments, and that may help for that particular argument, but the moment they leave us, then we normal specialists are still exposed to the original problem.

An entirely different kind of solution instead suggests itself: aggressive compartmentalization. Here, I think there is something significant we can learn from the field of computer programming, and especially the way that computer programmers have learned to manage the connections between software products made by different teams and different companies. When learning programming, one learns to program particular sequences of instructions — called subroutines or functions — that will be called from different places and times. These sequences can be triggered in various ways. For example, one might have a subroutine designed to create a digital image of Mario on a screen. The main program will trigger this subroutine over and over again, each time sending it a different set of input variables indicating where on the screen Mario should appear. These modules are called 'subroutines' when they are simply sequences of instructions, and 'functions' when they are sequences of instructions that also return information to the main program. For example, many programs use an "Add" function which takes any number of input variables and then returns, as output, their total. These functions often have considerable internal complexity. At first, the naive programmer is tempted to optimize their program by letting the other parts of their program make direct calls and modifications into the inner workings of particular functions, and let particular functions look for all sorts of particular variables from the larger program and from other functions. But the naive programmer quickly learns that this degree of functional transparency leads to chaos. Different programmers are always tinkering with the particulars of different functions and subroutines. We are taught, instead, the principle of modularity. In modular programming, one tries to make subroutines entirely independent

18

black boxes: a fully modular subroutine will take an input and generate an output, but everything that happens in between is inaccessible to anything outside of that particular module. Part of the expectation of modular programming is that a particular module may end up being called in all sorts of places and for all sorts of reasons that the programmer never had in mind when they originally created it. Modular programming becomes particularly important the more different programmers' programs need to interact. As soon as one has multiple programmers on a development team, or programs interacting from different development team, modularity becomes vital. Think again of that Mario subroutine. The main program fires off many other subroutines and functions, which will tell the main program where Mario is, what his vector is, and what the player's last input was. The main program would collect these variables and pass them to a Mario image-generating function. And that image-generating function would take those location variables as input, process them inside its black box, output a package of image data, and send that onwards to another subroutine designed to communicate with a specifically physical display device – like a TV screen, computer monitor, or portable game-playing screen. The final-stage programs that connect to actual specific physical output devices are called 'drivers'. But, since programs are run on different devices, they must be compatible with a wide range of drivers. Usually, a game development team will write their program so that it might hook up with any number of display drivers, for any number of display types. The game development team actually usually has no idea what's going on in the specific display drivers; in fact, the assumption is that drivers will change and update many times over the lifetime of their game, as the physical technology changes. So, game programmers and driver programmers can't rely on specific details of the other programs. Their modules must only exchange a small number of specific variables, designed for that exchange. The prime directive of modularization is the

isolation of individual functions, and the stabilization of any information exchanged across the interface between functions (Boudreau, Tulach and Wielenga 2007: 9-19). Another way to put it is that each module should depend as little as possible on the particulars of any other module. It should take an input, and then work independently to create the proper output, and other modules should look only to that particular output, and not otherwise peer inside the black box.

What might this look like applied to the broader problem of hyper-specialization? Here's an example: suppose you have created a better blood test for dopamine. All sorts of fields have a use for that blood test. Here are two options: first, you could translate — you could try to explain how the blood test works in terms that any scientist could understand, as carefully as possible, and then let it loose into the wider world. This process is complicated. The initial run of the blood through the analytic instrument yields a vast amount of data, which needs to be processed exactingly to yield a particular blood result. Then, animal researchers, medical researchers, biologists, and the like could try to make it work for themselves. This has the advantage of giving them access, not only to the dopamine data, but all the other complex instrumental data. Or, second, you compartmentalize and make it into a black box. You could set up a laboratory which performs this blood test. The laboratory asks blood samples to be collected in a specific way, processes those samples on-site and sends back a measurement of the dopamine level, and nothing else.

The first method seems good-hearted, intellectually generous, and hopeful in terms of supporting direct autonomy. Of course, the first method also invites hideous problems of cross-disciplinary defeater management. The second method, though perhaps intuitively repulsive to some, is an attempt to control for defeaters by *narrowing* the inputs and the outputs. The intent is to leave the defeasible innards of to the appropriate experts, and work

to manage the defeasibility condition of the exports, by sharply limiting what is exported. One might think that the second method is intellectually miserly. But another way to put it is that the second method is more intellectually humble. It recognizes the problem of cross-defeater management and recognizes that different experts in different fields simply to do not have the sensitivity to fully apply the arguments and methods of other fields.

The methods of modularity and isolation aren't a guaranteed solution to the problem of cross-disciplinary defeater management. But they do narrow the space of worries. Return to that blood test. If we opened up the details of that blood test to all sorts of fields, then we would have to worry about defeaters that might emerge at every step — of processing the blood sample, of submitting it to the instrument, of analyzing the data properly. But if we modularize, we only need to worry about and manage the defeaters at two points of contact: input and output. Here's how that might look: for anybody who orders the test, they receive meticulous instructions about how exactly to gather the blood sample. When the test results come out, they come with a long list of qualifications about what might have gone wrong. This doesn't solve the problem of defeater management, but it radically decreases the number of defeaters we need to manage.

When considering the problem of intellectual distance, we might have thought that the right thing to do at the interface between the fields was open the doors wider and let more understanding through. But the problem of cross-defeater management suggests the opposite: that we should narrow the opening between fields. Furthermore, it suggests that Millgram's army of philosophers should concentrate, not on explaining the innards of one field to another field, but rather on ensuring the clarity and stability of any information exchanged across the interfaces between fields, and on otherwise isolating the inner workings of one field across another.

Of course, radical modularity is not the only thing we can strive for. Let's return to the blood case. As it turns out, sometimes blood test developers need to export blood tests. A hospital may perform so much of a particular blood test that they wish to perform it in-house. If it's economically feasible, blood test developers will actually create literal black boxes — machines that can be sold to hospitals and, with minimal training, will take a blood sample and spit out a result. But we could also attempt to break out of modularity by double-training an individual. A zoologist who needed to adapt that blood test to gorillas might temporarily apprentice themselves to the blood test developer in order to learn the details of the blood test and optimize it for their needs. But this requires a massive expenditure of human time and effort — months if not years. Modules can be customized and custom-fit, but, as elsewhere in life, custom work is mighty expensive. We could also do this with whole fields; as Millgram suggests, we can take individuals and train them in two fields simultaneously, and those individuals can tend to the connection between those two fields. But this is likely only worthwhile when the two fields are deeply and frequently connected. For example, the relationship between aeronautics and the chemical engineering involved in jet engines likely has been tended to by a large number of specialists who are double-trained, as has the relationship between ethical theory and legal theory. To press the analogy further: in most of industrial life, we use modular parts. Sometimes, for a special purpose, we may custom-engineer a part, or custom-modify two parts for a better and more integrated fit. But that is time-consuming and expensive. Most of the times, we simply use modular parts and assemble them as needed.

In fact, one might start to suspect that modularity, while it depresses the traditional sort of direct autonomy of every individual piece of evidence, encourages yet another sort of autonomy: the ability to manage, for oneself, the overall shape of large systems of knowledge.

Let's call this *management autonomy*. Given that we are cognitively finite human beings, it seems like we have a choice: we can either know for ourselves all the details of a small system, but have little idea of where the inputs for that system come from or where the outputs for that system go — or we can step back and modularize, and get a glimpse of the whole. This permits us to autonomously consider the structural relationships of the various domains of knowledge. When human knowledge was small, it was possible to have all these forms of autonomy for oneself. But as human knowledge grows larger, it seems that we must choose between them. Direct autonomy, and its interest in transparency, is at odds with delegational autonomy's need to trust logical aliens and management autonomy's need to encapsulate fields in black boxes.

Conclusion

There is a tension between different potential reactions to the difficulties of hyper-specialization. First, to increase the familiar sort of intellectual autonomy, we might want to open the borders between the fields, translate the concepts. But this solution, though it increases our direct autonomy, will also exacerbate the problem of cross-disciplinary defeater management. To manage that second problem, we ought to narrow the interface between the fields, and compartmentalize instead of translate. That choice increases intellectual distance and decreases adaptability, and it requires greater trust and creates greater intellectual vulnerability, but increases our capacity for management autonomy. Furthermore, the drive to transparency presumes that we would wish to do all the thinking for ourselves, or employ somebody sufficiently like ourselves. But in some cases our

intellectual interests call us to place our trust in those significantly unlike ourselves. In that case, we want the capacity for delegational autonomy.

Are these forms of intellectual autonomy always going to stand in tension? Perhaps, in some sense, we can move towards transparency and modularity at once. Again, return to the computer programming example. It is possible for us to be simultaneously transparent and modularized, in at least once sense. Large swathes of the computer programming world are open-source — what's inside each program is available to anybody else, as source code. Any programmer can see what's inside any open source black box, and lift out the code and tinker with it. But, that's a matter of source code. When the programs are actually running, they behave as modules. If we export this analogy to the academic world, what this looks like might be a distinction between reporting and use. For example, it's perfectly fine for our blood test designer to publish her research and let other experts look under the hood. What functional modularity demands is that when, say, specialist in cell cancer wants to use that blood test, she ought to order the modularized blood test rather than try to adapt the original research for her ends. If we are worried about cross-disciplinary defeater management, then in actual use, modularity trumps transparency. If I am not a blood testing expert, then I may have a go at that research data in my own spare time, but when push comes to shove, I need to order the blood testing package and trust it, and defer to that trust, over my own shaky attempt at thinking through how that test works.

This may require that we abandon the familiar desire for complete direct intellectual autonomy and its concomitant fetishization of complete personal understanding and control. It might require, instead, trading off direct autonomy for trust in others, and for the capacity to manage large scale structures. What we're starting to uncover here is the fact that intellectual autonomy fragments under the pressure of the increasing size of human

knowledge. Or, perhaps, it was fragmented all along, and that fragmentation was hidden due

to the relatively small size of human knowledge. But as the amount of human knowledge

increases, the forms of intellectual autonomy separate, and come into tension, and we must

sometimes choose between them.

Bibliography

Baier, Annette, 1986, "Trust and anti-trust," in *Ethics,* 96, 2: 231-260.
Boudreau, Tim, Jaroslav Tulach, and Geertjan Wielenga, 2007, *Rich Client Programming: Plugging Into the Netbeans Platform*, Prentice Hall Press, Upper Saddle River.
Cholbi, Michael, 2007, "Moral expertise and the credentials problem," in *Ethical Theory and Moral Practice*, 10, 4: 323-334.
Hardwig, John, 1985, "Epistemic dependence," in *The Journal of Philosophy*, 82, 7: 335-349.
Kitcher, Philip, 1993, *The Advancement of Science*, Oxford University Press, New York.
LaBarge, Scott, 1997, "Socrates and the recognition of experts," in *Apeiron: A Journal for Ancient Philosophy and Science*, 30, 4: 51-62.
McGrath, Sarah, 2011, "Skepticism about moral expertise as a puzzle for moral realism," in *Journal of Philosophy*, 108, 3: 111-137.
Miller, Daphne, 2008, *The Jungle Effect*, Harper Collins, New York.
Millgram, Elijah, 2015, *The Great Endarkenment: Philosophy for An Age of Hyperspecialization*, Oxford University Press, New York.
Nguyen, C. Thi, 2018, "Cognitive islands and runaway echo chambers: problems for epistemic dependence on experts," in *Synthese*, https://doi.org/10.1007/s11229-018-1692-0
Nguyen, C. Thi, 2010, "Autonomy, understanding, and moral disagreement," in *Philosophical Topics,* 38, 2: 111-129.
Nickel, Philip, 2001, "Moral testimony and its authority," *Ethical Theory and Moral Practice*, 4 ,3: 253-266.
Nozick, Robert, 1990, *Examined Life: Philosophical Meditations*, Simon and Schuster, New York.
Swanwick, Michael, 2011, *Stations of the Tide*, Macmillan, New York.