Introspection, What?

Eric Schwitzgebel
Department of Philosophy
University of California at Riverside
Riverside, CA 92521

eschwitz at domain- ucr.edu

May 12, 2011

Introspection, What?

<u>1. Thesis and Alternative Views.</u>

My thesis is: Introspection is not a single process but a plurality of processes. It's a plurality both *within* and *between* cases: Most individual introspective judgments arise from a plurality of processes (that's the within-case claim), and the collection of processes issuing in introspective judgments differs from case to case (that's the between-case claim). Introspection is not the operation of a single cognitive mechanism or small collection of mechanisms. Introspective judgments arise from a shifting confluence of many processes, recruited opportunistically.

The following analogy might be helpful. Suppose you're at a psychology conference or a high school science fair and you're trying to quickly take in a poster. You are not equipped with a dedicated faculty of *poster-taking-in*. Rather, you opportunistically deploy a variety of processes with the aim of getting the gist of the poster: You look at the poster – or perhaps only listen to a recital portions of it, if you're in the mood or visually impaired – you attend to what the poster's author is saying about it; you follow out implications, charitably rejecting some interpretations of the poster's content as too obviously foolish; you think about what it makes sense to claim given the social and scientific context and other work by the author or the author's advisor, if you know any; you pose questions and assess the author's responses both for overt content and for emotional flavor. Although the cognitive systems involved range widely and are not dedicated just to taking in posters, not just any activity counts as taking in a poster – one's judgments about the poster must aim to reflect a certain kind of sensitivity to its contents. Likewise for introspection, I will suggest: The cognitive activities range widely and vary

between cases – that is the main claim I will defend – and yet, as I will suggest near the end of this essay, it wouldn't be natural to call a judgment introspective if it wasn't formed with the aim or intention of reflecting a certain kind of sensitivity to the target mental state.

As far as I can tell, no previous philosopher or psychologist has defended both within-case and between-case pluralism about introspection. Although defenders of the view that introspection (or "reflection" or "inner sense") resembles sensory perception of the outside world could have developed this view in a pluralist direction, historically they have not done so (e.g., Locke 1690/1975; Kant 1781/1787/1997; Wundt 1888, 1896/1897). Some philosophers and psychologists have distinguished between two types of processes that can lead to introspective or quasi-introspective judgments, but two is not many, and often the point of such distinctions is to isolate a unitary target process of interest (e.g., Brentano 1874/1973; Wundt 1888; Russell 1912; Nichols and Stich 2003).

Contemporary philosophers tend to adopt one of two perspectives on introspection and self-knowledge. One approach characterizes introspection (or "self-awareness") as the operation of a mental self-scanning or self-monitoring process (e.g., Armstrong 1968; Lycan 1996; Nichols and Stich 2003; Goldman 2006). Typically, this process is characterized as involving a single fairly simple detection or monitoring mechanism or family of closely related, simple monitoring mechanisms. Small complexities can plausibly be added: Nichols and Stich (2003) couple the monitoring mechanism with a second, sometimes-competing, sometimes-cooperating mechanism, involving the application of a general "theory of mind"; Goldman (2006) couples monitoring with a capacity to "redeploy" representational contents and to translate representations from one type of mental code into another. But the addition of such complexities doesn't constitute broad pluralism.

A second approach emphasizes mechanisms or procedures other than self-monitoring as the primary ground of self-knowledge or of the privileged self-ascription of mental states. Broadly speaking, these approaches fall into five classes:

- *Self-fulfillment:* Self-ascriptions might involve embedding a target content within a self-ascriptive content in a self-fulfilling way. For example, "I am thinking of a banana" might be automatically true because the thought that I am thinking of a banana contains within it the thought of a banana. (See Burge 1988, 1996; Heil 1988; Gertler 2001, this volume; Papineau 2002; Chalmers 2003; Horgan and Kriegel 2007; Horgan this volume; and for a somewhat different version, Shoemaker 1996, this volume; arguably this maneuver goes back to Descartes 1641/1984.)

- *Self-shaping:* Self-ascriptions might tend to be true because in making a self-ascription one is committing to a certain way of thinking or acting, presently or in the future, in accord with one's self-ascription. For example, the self-ascriptive thought "I hate laundromats" might be casually influential in creating or sustaining hatred of laundromats. (See Moran 2001; McGeer and Pettit 2002; McGeer 2008.)

- *Self-expression:* Self-ascriptions like "that hurts!" or "I don't wanna" might essentially be complicated ways of wincing or frowning, or of saying "ow!" or "that stinks!" – that is, they might be self-ascriptive linguistic variants of ordinary spontaneous expressions, which require no prior self-scanning. (See Wittgenstein 1953; Bar-On 2004; Gordon 2007.)

- *Direct inference:* Self-ascriptions might be derived inferentially, or quasi-inferentially, directly from judgments about the outside world, requiring no more introspective self-scanning than does ordinary inference. For example, from *P*, just as one might

straightaway conclude *P or Q*, without introspecting one's judgment that *P* (presumably

not all inference requires introspection), similarly one might straightaway conclude *I*

believe that P. (See Dretske 1995; Tye 2000; Byrne 2005, this volume.)

- *Theoretical inference:* Self-ascriptions might be largely grounded in observations of

   one's own behavior, combined with the theories of folk psychology. For example, from

   the fact that you often drive across town to get Thai food, you might infer that you like

   Thai food. (See Bem 1967; Nisbett and Wilson 1977; Gopnik 1993a&b.)

For a fuller review of these approaches to self-knowledge, see Schwitzgebel 2010. Advocates of

such views tend not to embrace broad pluralism, though they may emphasize a single competing

process.

Prinz (2004, 2007) and Hill (2009, 2011) both defend explicitly pluralist views of

introspection, but their pluralism appears to be only between-case: Although they suggest that

many different kinds of cognitive mechanism can yield introspective knowledge, they tend to

portray individual introspective judgments as issuing each from the operation of a single

mechanism. (Prinz also seems to regard as "introspective" many processes that most

philosophers and psychologists would not ordinarily regard as introspective, such as, apparently,

ordinary simple recall; see also my discussion of Hill in Schwitzgebel 2011b.)

The blind men have each, it seems to me, nicely described a piece or two of the elephant.

But none have adequately displayed the pieces' integration into a messy, moving organism.

2. Examples.

   2.1. Visual experience. I look out the window and reach the judgment not only that

there's a tree outside but also that I'm having a visual experience of that tree. I have greenish

visual experience of the leaves, and the tree's spreading branches seem to dwarf the mountain in the background. It has just rained, and in the re-emerging sun, the tree sparkles strikingly. Focusing my gaze on the rightmost branches, I notice a fluttering indistinctness in my experience of the left side of the tree. I cross my eyes, thinking it might make the tree double, but instead the tree only swims around my visual field, blurring and flattening. So, how do I know all this about my visual experience?

Let's begin here: You, standing next to me, seeing me look attentively out the window, might reach some of the same conclusions about me. Minimally, you can safely guess that I'm having greenish visual experience of some sort. After all, you know (a.) that I am looking at a green thing in good conditions, and (b.) (let's suppose) that I'm not colorblind. Now of course I, too, know (a) and (b) about myself. Might (a) and (b) be part of my grounds in thinking that I'm having a greenish visual experience? They seem unlikely to be the *sole* grounds of my judgment – presumably, if I were to have gone suddenly and unexpectedly colorblind (through, say, the secret action of a prankster neuroscientist) I would notice that I'm not in fact having greenish visual experience, despite my looking attentively at what I know to be a green object in plain view. But even if (a) and (b) aren't the sole grounds of my judgment, it seems reasonable for me to allow my knowledge of outward objects and my own capacities to play some causal and justificatory role in my knowledge of my visual experience. If I know that I'm looking at an evenly painted white surface, I might more naturally reach the judgment that I'm having a visual experience of even whiteness than if I know that I'm looking at a surface with a gradual shift in color. If I know that the burrito I'm biting into has cheese in it, I might be naturally and justifiably primed to judge that it tastes "cheesy". If I see you move behind me with a red-hot poker and then suddenly I feel a startling touch on my neck, I might swiftly and readily judge

that I'm feeling heat and pain, not coolness, even if you have actually touched me with an ice cube. (Maybe I do, in fact, for a split second feel heat and pain, or maybe not; what's at issue right now isn't that, but only the contribution my expectations do and should make to my judgment about my experience.) I know the tree has leaves; I know it has just rained; I know what trees in general are like, and what the scene from that window is normally like. All this knowledge influences, it seems plausible to suppose, not only my experience but also my expectations about my experience, my readiness to make certain judgments about my experience, and thus those experiential judgments themselves. If my experience is other than expected, I am called up short; I hesitate; it takes a little time, perhaps, and some reconfirmation before I come around, if indeed I do come around.

When I try crossing my eyes and don't receive, or don't think I receive, the expected double vision of the tree, I react in part by wondering whether I have really succeeded in crossing my eyes; so I try again, wiggling my eyes in various ways, using some combination of motor intentions, proprioceptive feedback, and visual feedback to assess the state of my eyes. I wonder, too, whether there might be a double image of the tree that I'm failing, at least momentarily, to notice (Helmholtz 1856/1909/1962; Titchener 1910; Schwitzgebel 2011a, ch. 2). When I hold my eyes fixed on the tree's rightmost branches and fail to discern the details of the leftmost, my failure in that seemingly outward visual task is part of the basis of my judgment that my visual experience is indistinct away from the point of fixation. My sense of my visual experience is probably shaped, too, by culturally available metaphors, especially painting, photography, and movies, which might draw me toward thinking of my experience as in some way flat like a painting or as possessing distinctness of shape and color well into the periphery (Noë 2004; Schwitzgebel 2011a, ch. 2). Why am I inclined to think of the tree as dwarfing the

mountain? Does this have to do with the projective size each would have on flat media, or the visual angle subtended?   Is there also a sense in which the mountain looks much bigger than the tree? How stable and well-grounded and culturally variable are such judgments about the experience of size? The tree sparkles in the sunlight in a way I find striking, and my judgment that this aspect of the scene is striking is partly phenomenological, partly cognitive or aesthetic – a judgment that probably interacts loopingly with my knowledge of the environment, my knowledge of my visual experience of the environment, and my knowledge of other aspects of my reaction to the scene. I hear myself speak, inwardly or outwardly, I shift my gaze, I shift my attention without shifting my gaze, and those processes, too, influence both my visual experience and my apprehension of my visual experience.

My judgments about my experience, then, are influenced by at least: my expectations about my experience, my knowledge of the outward environment, my knowledge of what I can and cannot discern, culturally available metaphors and general theories about visual experience, and my knowledge of other aspects of my psychology, in temporally entangled loops. Is there, embedded within this tangle, a distinct, genuinely introspective process, separable at least in principle from any non-introspective influences upon the various emerging judgments? I feel the pull of that idea. The arguments in some of the subsequent essays in this volume appear to turn on the possibility of isolating, in principle, a purely introspective process from amid such noise (especially Gertler this volume; Sosa this volume; Zimmerman this volume). My suggestion in this essay, however, is that it is best to resist treating introspection as distinctive and isolatable. There is no important, cognitively distinct process that is the process of introspection, pure.

The view that introspection of visual experience is a process distinct from the processes of visual perception, when that view is combined with a broadly self-monitoring approach to

introspection, appears to invite the following cognitive model: First there is a process of visual perception, and then afterwards begins the process of introspecting one's perceptual experience. Maybe the first process, the perceptual process, continues while the introspective process works at a delay upon its results, always a stage or more behind. However, I doubt this is the best way to conceive of the cognitive processes involved in my example of looking at the tree. It's more useful, I suspect, to treat the ordinary perceptual processes of vision in that case as part of, or as overlapping with, the introspective processes that shape my judgments about my experience. My visually obtained and constantly updated knowledge of the objects around me is a crucial part of the cause and grounds of my judgments about my visual experience of those same objects. So is the process of trying and failing to visually discern properties of the world. If perception is a complex looping process involving activities of the body such as the movement of fingers and eyes (e.g., Hurley 1998; Noë 2004), so too, I suggest, is introspection in the example above: My activity of holding my eyes still and attempting to discern the shape of the leaves in the periphery, my activity of trying to determine if I have successfully crossed my eyes, my looking around, my recruitment of general knowledge and knowledge specific to the situation, are all part of a multifaceted project that it is artificial to try to divide into introspective and non-introspective pieces.

Here is another phenomenon that strains against the idea that introspection is a cognitively distinct process sharply separable from the processes of outward perception: Judgments about sensory experience can easily collapse into judgments about the outside world with no crisp border between; and the two sorts of judgments, in such cases, are often seemingly driven by virtually identical cognitive processes. So, for example, if asked, for each of a series of stimuli, to report on one's visual experience of the color of the stimuli, one might first say

"green", then "red", then "green again", with the explicit intention of reporting only on one's visual experience, that is, on a piece of phenomenology rather than on properties of the outward stimuli. But after settling into the monotony of the task, it is quite natural to slip absent-mindedly into expressing, instead, one's judgment about the outward stimuli themselves – the colors of the material objects – especially if there's no reason to doubt that one's perception is veridical. Such slipping was my frequent experience in reading Titchener's famous manual of introspective training and attempting to replicate some of his exercises; and Titchener felt it necessary to repeatedly warn aspiring introspective experts against such "stimulus error" or "R-error" (Titchener 1901-1905; see also Boring 1921; Schwitzgebel 2005; 2011a, ch. 5). Where one suspects illusion, "is green" and "looks green" express very different judgments; where one does not suspect illusion, they can blur into each other. Despite their different truth conditions, the two sorts of judgment – one about the stimulus object, one about the experience of that object – are often difficult to pull apart psychologically. We generally use the same terms to express both the objectual and the introspective judgment (e.g., "green" for both the property of the object and the property of the experience the object produces in me), and often, it seems, there is no discrete fact of the matter which of the two judgments I am making or whether I am making both simultaneously. We gradually, insensibly traverse the distinction between introspective and non-introspective judgment. In such cases, introspection might be best regarded as perception with a twist or with a slightly different aim that can be half forgotten. The processes of perception, then, would be *part* of the process of introspection.

    2.2. Emotion. I think about what, if anything, I am emotionally experiencing right now. I notice, first, that my lips are pursed, and I relax them; I notice some tension in my chest. But then I think to myself that emotional experience is not, or might not be, entirely bodily. In fact,

it seems a little odd that I should leap straight to bodily self-apprehension in thinking about my emotion. Do I usually do this when I reflect on my emotional experience? Some kind of negative affect is present – perhaps I'm tense about writing this essay? A visual image of a blank word-processing screen has come before my mind. But I had been looking forward all day to finally having a chance to write! As I think about the little remaining time to write this essay, I seem to become more unsettled. I am tense, I decide, about the looming deadline. I find my lips pursed again and rub them with my left hand. There is a bit of an odd feeling in my cheeks, but I don't know if it is associated with the emotion. Being tense about the deadline doesn't seem like the only thing that is going on with me emotionally right now – but what more there is I can't quite put a finger on. I find myself listening to the freeway traffic in the distance, calming myself a bit with eyes closed and head in hands, and then I imagine how I would look to someone viewing me from behind. I think that thinking about this particular introspective task is worsening my mood, making me tenser and maybe almost angry. I would have liked a happier example, one that better displays the sunny disposition and amiable character I believe myself to have. Perhaps, partly, I am distressed at the negativity of this example, and that distress is further reinforcing the negativity. Maybe, too, there is some self-shaping involved: Maybe what I was really experiencing was a relatively undifferentiated negativity or tension, induced partly by over-caffeination, and it became concretized, in part, as deadline anxiety largely as a result of my entertaining that hypothesis.

My introspective – or seemingly introspective – assessment of my emotional state seems to flow, again, from multiple sources, including at least my knowledge of my social environment (I am writing an essay under deadline pressure), proprioceptive knowledge of my body (my lips are pursed, my throat tense), and knowledge of my thoughts and imagery (knowledge which is,

in its turn, presumably also at least partly introspective); perhaps, too, there was some relatively direct causal influence from my emotional phenomenology to my judgment about it. (See Section 4 for a discussion of whether such direct causal influences might be what introspection really is.) The process was temporally extended, and I noted how things seemed to shift as my mind moved from topic to topic. It's almost like I was pulling together pieces of a story. Jakob Hohwy (forthcoming) has argued that introspective knowledge, like sensory knowledge, tends to be exploratory: We have initial expectations or hypotheses, and on the basis of them our cognitive systems make predictions about how things will change (or remain the same) over time, across various conditions. To test these hypotheses we often act to alter those conditions and we adjust our assessment of the probabilities according to whether our implicit or explicit expectations of change are confirmed or violated. In sensation, we might move the object or our heads, or we might tap on the object. In the present example, I fix my attention upon the hypothesized source of anxiety, the looming deadline, and note what seems to change in me as a result. More subtly, perhaps, I introduce and track proprioceptive changes when I rub my lips. On this view, the motion of my hand against my lips and the act of bringing the deadline vividly to mind are parts of, and not simply preconditions of, the exploratory introspective process. If I pursue the wrong contingencies or draw the wrong conclusions from them, then I have failed not only to set up the preconditions of introspection in the way I would have liked, but I have failed in the introspective task itself. For example, I err introspectively if sinusitis is the cause of my cheeks' discomfort and yet the discomfort's failure to recede when I turn my mind away from the looming deadline contributes to a false impression that I'm anxious about more than just the deadline. As in the visual case, bodily, perceptual, and cognitive processes that are not intrinsically introspective can become part of or overlap with the introspective process. On the

approach advocated in this paragraph, we cannot say that here on the one hand are the outward bodily and perceptual processes and here on the other hand is the purely introspective process.

Shifting example, my knowledge that I'm afraid of the rattlesnake or (perhaps differently) that I'm *feeling* afraid of it, might derive in part from my general knowledge that rattlesnakes are dangerous and my visual knowledge that one is only three feet away; it might derive in part from my proprioceptive and visual knowledge that I've just flinched, from my knowledge that I felt a tingling surge of what I would call adrenaline, from my sense that I have the impulse to run; it might derive in part from my knowledge that I just uttered an expletive, either in inner or outer speech, from an awareness that I'm imagining the snake biting me, from a kind of numb paralysis I feel; I might have an impulse to say to my hiking partner "I'm terrified of that snake", which I do or do not disinhibit. Uttering those words, in fact, may help make them true – or at least congeal my emotion and give it specific shape. Probably too, there are some low-level neural or cognitive connections that operate in none of these ways but work more directly from my fear to enhance the likelihood that I will judge myself afraid. Note my appeal here to several of the processes described in Section 1: self-detection, self-shaping, self-expression, theoretical inference.

2.3. Imagery. In the morning on the way to work, I blasted a tune on the car stereo, Sonic Youth's "Kool Thing". Now it's two in the afternoon and I notice that that tune has been running through my head. It has been running for at least three seconds, I think, maybe much longer, and not for the first time today. As I reflect, the tune seems to sharpen or become more vivid. It seems that I can choose to emphasize the vocals or the guitars, and I think about the extent to which I can imagine both the vocals and guitars simultaneously. I conclude that I can

do so, especially if I nod my head in rhythm and do something that feels like using my mouth and voice to track the lead guitar line (though no noticeable sound issues from my mouth).

Imagery is so much under our immediate control that concurrent introspective judgments about it seem bound, in most cases, to be supported to some extent by self-shaping, that is, by the process of controlling my mental life in such a way that it conforms to my judgments about it (rather than simply the other way around). Imagery judgments might even be self-fulfilling, if the target image can be a part of the self-attributive judgment about it. When I judge that I'm visually imagining my mother's face or hearing the chorus of "Kool Thing" in my head, I am partly working to make these self-attributions true as I reach them. However, it seems unlikely that I can make just any judgment about my imagistic phenomenology true simply by willing it to be so: If I judge that I am visually imagining the Taj Mahal with every arch and spire simultaneously well defined, or that I am imagining, simultaneously, the vocals, bass line, drums, and both guitars, or that I am visually imagining a triangle that is neither equilateral, isosceles, nor scalene but somehow all and none of these at once (Locke 1690/1975; contra Berkeley 1710/1965), I might be wrong. And hopefully if I am wrong, something in me – some influence, direct or indirect, from the imagery experience itself – will lead me to refrain from the attribution, or cancel it, or at least hesitate and feel uncertain. When I'm trying to determine if I can imagine the lead guitar and vocals at the same time, it seems that I'm not only creating or sustaining the imagery but also checking to see if I have successfully created it as intended. Self-regulative feedback is integral to bodily action, except in the swiftest ballistic movements; we might think of imagery creation similarly.

Of course, self-shaping and self-fulfillment can't explain knowledge of very recently past imagery, since self-shaping and self-fulfillment are necessarily present- or future-oriented. And

often, too, there is little environmental or inferential basis for judgments about the contents of one's recently past imagery; nor does it seem that we can directly, and non-introspectively, self-express past and gone mental states in the way, perhaps, that we can directly and non-introspectively burst out with a "that hurts!" or "I don't wanna". Thus, recently past imagery is a case where relatively direct causal influences from conscious experience to one's judgments about it are most evident – thus revealing the incompleteness of any account of self-knowledge limited to the five non-self-monitoring procedures mentioned in Section 1: self-fulfillment, self-shaping, self-expression, direct inference, and theoretical inference. (I am assuming here that the operation of those five methods is not continuous or very frequent; otherwise, some combination of those five methods plus memory of their outputs might explain self-knowledge of recently past imagery.) These relatively direct influences from immediately past imagery might take a variety of possible forms: The influences might be mediated by short-term memory, or iconic memory, or a looping process; or they might involve fading activation or the normal temporal course of a feed-forward causal brain process; or they might reflect a partial temporal overlap between cognitive processes. The empirical question is open, but here as elsewhere I'll bet on multiplicity.

In noticing "Kool Thing" running through my head, it seems likely, then, that I'm partly shaping it as it transpires, to conform with my judgments about it, and partly exhibiting some relatively direct sensitivity to the experience that is thereby created. Plausibly, too, as in the visual experience and emotional cases described above, my judgment about my experience draws upon general knowledge that makes various experiences or features of experiences seem more or less likely. That knowledge might include: what would be a plausible memory image, given what I know about the band's usual instrumentation, about that style of music, and about

that particular song; my opinions about imagery in general (which are likely to be partly culturally conditioned); and my possibly accurate or possibly distorted opinions about my own imagery capacities (see Schwitzgebel 2011a, ch. 3). Perhaps, too, I am simply apt to burst into song as a way of expressing, and simultaneously concretizing, my knowledge of my imagery experience.

3. The Boxology of Introspection.

It's often helpful for cognitive scientists modeling psychological processes to describe the mind's functional architecture using boxes and arrows, with the boxes indicating various functionally discrete processes or systems and the arrows indicating the causal or functional relationships among those discrete processes or systems. Figure 1 expresses my view of introspection, using the "boxology" of cognitive science. The model in that figure may be contrasted, for example, with the boxological models on pages 162 and 165 of Nichols and Stich 2003, which feature tidy arrows in and out of the Belief Box, through a Monitoring Mechanism, a Percept-to-Belief Mediator, and a Theory of Mind Information store. You might also notice a resemblance between my model in Figure 1 on the next page and recent boxological models of visual processing, if the latter are squinted at.

Figure 1: The boxology of introspection
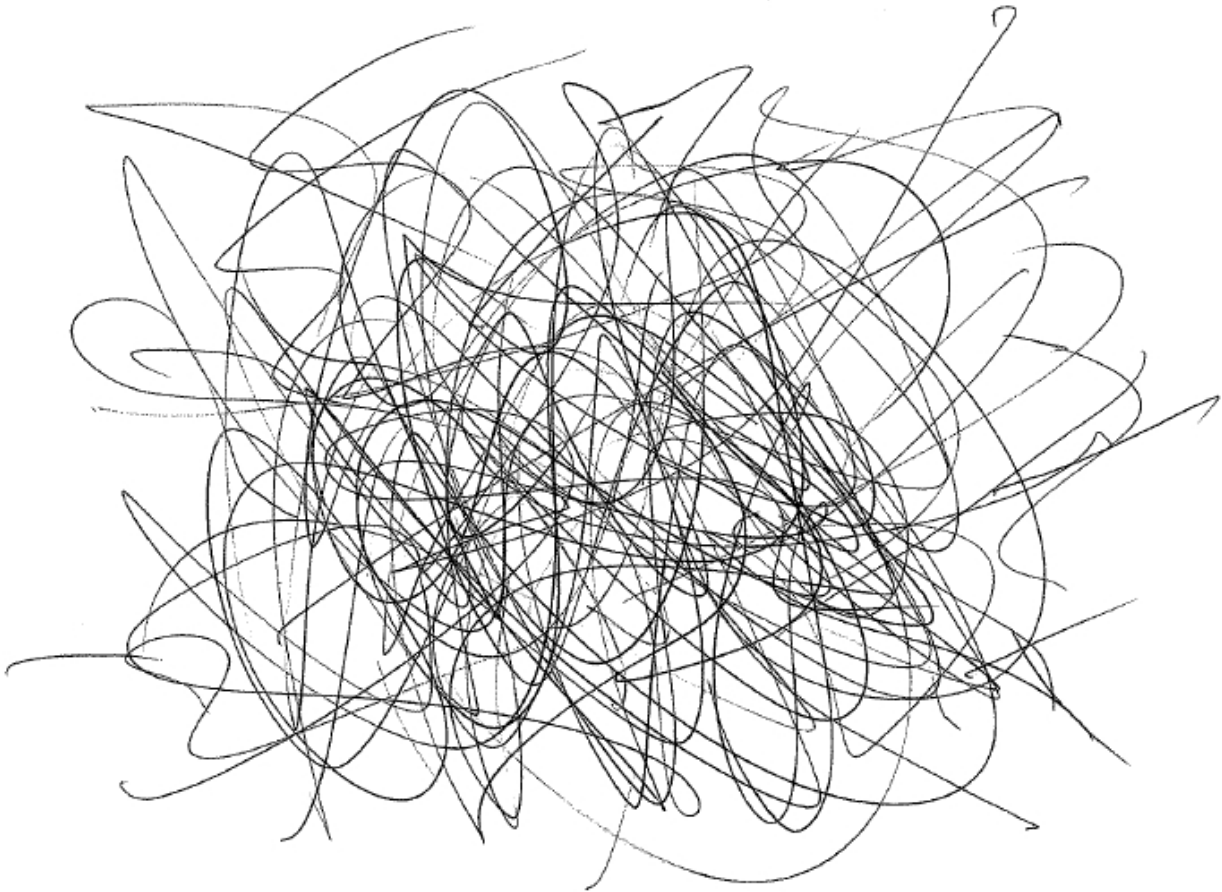
Three broad considerations favor this boxology of introspection. First, each of the methods of arriving at self-knowledge described in Section 1 seems appropriate to some cases, and the various methods appear to have considerable potential to compete or co-operate in individual instances; and, furthermore, as Prinz (2004) and Hill (2009) argue, it seems unlikely even on a pure scanning view that there would be a single type of scanning mechanism for all possible target states. These considerations suggest substantial between-case pluralism, at least. Second, as I hope the examples of Section 2 illustrated, it seems plausible that in many cases of apparently introspective self-knowledge a wide variety of resources and capacities are brought to bear on the judgment, varying both within and between cases. And third, the more neuroscientists discover about the massive interconnection of the brain, the more it seems architecturally likely that, generally speaking, people's conscious judgments will draw upon a large variety of influences, from the short and direct to the loaded and circuitous.

It's worth noting, perhaps, that similar considerations recommend a similar boxology for other broad, person-level cognitive processes, like memory, visual perception, and decision. I support that generalization of the diagram (and I briefly discuss the complex influences on memory judgments in Hurlburt and Schwitzgebel 2011), with three qualifications: First, I don't intend to deny entirely the existence of simple or functionally isolated cognitive processes – perhaps some of the processes operating early in the visual stream are approximately functionally isolated, for example. But such processes, if they exist, are unlike introspection, memory, or visual perception; they are not the broad types of cognitive processes recognized by the terms of folk psychology and capable by themselves of generating conscious judgments. Second, empirical investigation can weave a bit of order out of the chaos, allowing us to reify features of the swirl into a large variety of interacting sub-processes, partially isolatable, at least

as an approximation. This has already occurred to some extent for memory and visual perception, and it may start to occur for introspection as psychology starts more seriously to contemplate its mechanisms. However – and this is my third qualification – there are well-established partial deficits of memory and vision that suggest a certain degree of functional separability among sub-processes; there is currently no parallel taxonomy of partial introspective deficits – no clear pattern, for example, of functional double dissociations among introspective sub-processes (pace Nichols and Stich 2003; see Carruthers forthcoming).

4. Introspection, What?

I doubt that we can draw sharp lines through this snarl, cleanly isolating some genuinely introspective process from related, adjoining, and overlapping processes. What we have, or seem to have, is a cognitive confluence of crazy spaghetti, with aspects of self-detection, self-shaping, self-fulfillment, spontaneous expression, priming and association, categorical assumptions, outward perception, memory, inference, hypothesis testing, bodily activity, and who only knows what else, all feeding into our judgments about current states of mind. To attempt to isolate a piece of this confluence as *the* introspective process – the one true introspective process, though influenced by, interfered with, supported by, launched or halted by, all the others – is, I suggest, like trying to find the one way in which a person makes her parenting decisions, the one cognitive process behind writing a philosophical essay, or (to return to the example from the beginning of the paper) the one cognitive process of taking in a science poster. The causes, the influences, the considerations, are too rich within most cases and too variable between cases for any but a radically pluralist account to do justice to the phenomena.

One might try to go subpersonal: If there is a cognitive subsystem with the task of keeping a bead on happenings in other parts of the mind, then perhaps that is the introspective system, even if it alone is not responsible for the judgments we arrive at? But surely there are many such systems, if there are any: The mind is thoroughly entangled and the different parts, and subparts of those parts, are designed to track and respond to goings-on elsewhere, at the micro-level as well as the macro-level, from relatively early to relatively late stages of processing, often beyond what we would normally consider to be our introspective ken. This kind of inter-system tracking seems hardly sufficient for introspection, at least in any ordinary sense of the term. What must be added to such processes to render their operation the operation of introspection proper? I suggest that they must get tangled up with the whole variety of processes that drive person-level conscious judgment.

One might attempt some sort of self-fulfilling content embedding story (as briefly described in Section 1, a view that seems recently to have gained momentum): Introspection involves loading target mental states into judgments about those very states: "I'm thinking about a hedgehog", "I'm experiencing [this]" (see, e.g., Gertler's and Horgan's contributions to this volume). We can reach infallible judgments in this way, perhaps (just as when we say "this sentence refers to itself" we necessarily speak the truth) – but that very infallibility shows that we have missed our target, or at least the target that I and probably most other people have in mind when we think about introspection: For introspection in practical use is *not* infallible; we don't always get it right in our introspective judgments about our emotional states, about the level of detail in our imagery, about the various features of our visual experiences, our pains, our inner speech. Elsewhere, I have argued that we err very often (Schwitzgebel 2011a, ch. 7), but the frequency of error isn't as much the present issue as the possibility of error. The kind of

introspection that matters to human affairs and to the methods of psychology and consciousness science is the kind of introspection that involves the fallible application of categories in explicit, conscious judgments – and thus involves the wide variety of resources we flexibly bring to bear in reaching such judgments.

But it would be a wacky sort of pluralism that counted every cognitive process as introspective, so let me conclude by suggesting some boundaries. The kinds of examples I have offered, and the kinds of cases I think most philosophers have in mind when they discuss introspection, are cases in which we arrive at explicit, conscious judgments about our own current or very recently past mental states. So perhaps we can say that we ought not regard as introspective any confluence of processes that fails to issue in that sort of judgment, or fails at least to be headed in the direction of issuing in that sort of judgment (though it may be cut short or collapse for some reason). And if the proper product of introspection is a conscious judgment, then we can also say, I think, that introspection consumes attention, on the assumption that forming a conscious judgment necessarily consumes attention.

I am inclined to recommend the following view: Introspection is the dedication of central cognitive resources, or attention, to the task of arriving at a judgment about one's current, or very recently past, conscious experience, using or attempting to use some capacities that are unique to the first-person case (like most of the capacities emphasized in the accounts in Section 1), with the aim or intention that one's judgment reflect some relatively direct sensitivity to the target state. It by no means follows from this characterization that introspection is a single or coherent process or the same set of processes every time. Now of course I can arrive at conscious, explicit judgments about my current or very recently past conscious experience without doing anything like what we would normally consider introspection: For example, I might read the outputs of a

neuroimaging machine, apply a general theory about how those outputs relate to consciousness, and (hypothetically at least) arrive at a judgment about my current conscious experience on that basis alone. Thus the characterization above requires that introspection involves the attempt to use capacities unique to the first-person case and that reflect a relatively direct sensitivity to the target state. Likewise, there seems something odd about calling a judgment introspective if it is *entirely* a matter of creating the target state in the course of self-ascribing it with no aim or intention that one's judgment reflect sensitivity to that state. One further consequence of that last condition is that self-attributions that pop to mind unbidden are not introspective – or rather, they are not introspective *unless* we are liberal about what counts as having the relevant aim or intention. Maybe we should be liberal. I prefer to leave the matter vague, allowing for in-between cases and stronger and weaker senses of "introspection". I also leave it vague what counts as "relatively direct". Attempting to specify too precisely the boundaries of introspection would require, I suspect, knifing more sharply through the spaghetti than the phenomena warrant.

Is this, then, really a "multiple realization" view of introspection? And if so, is it consequently just a variety of ordinary functionalism? In a way. But here's the twist: Just as the functionalist about pain denies that pain is a single type of physical process, because pain can be variously realized at the physical level, so also would I deny that introspection is a single type of cognitive process, since introspection can be variously realized at the cognitive level. Despite our ability to gesture at a class of cognitive activities we might call "introspective", no common cognitive core is shared by all and only introspective processes. To make that last point is, I think, just to restate (a modest form of) between-case pluralism about introspection. Within-case pluralism – at least as developed in Section 2 – adds the further thought that the processes

constituting any single introspective event will normally be, in large part, a combination of processes that exist primarily to serve non-introspective cognitive functions. If this view is functionalism, it isn't the type of cognitivist functionalism that treats diverse physical processes as nonetheless cognitively unified. Whatever unification there is, exists at a higher or different level of abstraction – perhaps only amid the rather vague abstractions of folk psychology.

A final clarification: My characterization of introspection limits the targets of introspection to conscious experiences. Now while the most central and uncontroversial examples of introspection – and all the examples I have used in this essay – take conscious experiences as their targets, philosophers often suggest that introspection can also take another important class of targets, to wit, attitudes, like belief and desire. Can't we also introspect those? I propose the following: If an attitude is consciously experienced, we can introspect it, and its availability as a target of introspection is already permitted by the characterization above as its stands. On the other hand, if an attitude is not consciously experienced, then it seems – just as a matter of empirical fact? – that we can learn about it only relatively indirectly, using roughly the same variety of tools we use to learn about other people's attitudes (though supplemented with a more direct knowledge of potentially related conscious states like inner speech, imagery, or emotional experience; see Carruthers forthcoming; also Ryle 1949; Goldman 1993, 2006; Hill 2009). Thus it would be misleading to say that we introspect non-conscious attitudes – misleading because it would suggest that we can discover them in part by deploying capacities and processes, or a certain range of capacities and processes, unique to the first-person case.

Philosophers have often characterized introspection as fundamentally epistemically superior to perception, cognitively or structurally simpler than perception, and perhaps also prior to perception and more foundational. If the picture I have sketched in this essay is correct, such

claims are all false.[1]

References:

Armstrong, David M, (1968). *A materialist theory of the mind*. London: Routledge.

Bar-On, Dorit (2004). *Speaking my mind*. Oxford: Oxford.

Bem, Daryl J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74,* 183–200.

Berkeley, George (1710/1965). *A treatise concerning the principles of human knowledge*. In *Principles, Dialogues, and Philosophical Correspondence*, Colin M. Turbayne (ed.). New York: Macmillan.

Boring, E.G. (1921). The stimulus error. *American Journal of Psychology, 32:* 449-471.

Brentano, Franz (1874/1973). *Psychology from an empirical standpoint*, Oskar Kraus and Linda L. McAlister (eds.) and Antos C. Rancurello et al. (trans.). London: Routledge.

Burge, Tyler (1988). Individualism and self-knowledge. *Journal of Philosophy*, 85: 649–663.

Burge, Tyler (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society, 96,* 91–116.

Byrne, Alex (2005). Introspection. *Philosophical Topics, 33,* 79–104.

Carruthers, Peter (forthcoming). *The opacity of mind.*

Chalmers, David J. (2003). The content and epistemology of phenomenal belief. In *Consciousness: New philosophical perspectives*, Quentin Smith and Aleksandar Jokic (eds.). Oxford: Oxford.

Descartes, René (1641/1984). *Meditations on first philosophy*, in *The philosophical writings of Descartes*, vol. 2, John Cottingham, Robert Stoothoff, and Dugald Murdoch (eds.). Cambridge: Cambridge.

Dretske, Fred (1995). *Naturalizing the mind*. Cambridge, MA: MIT.

Gertler, Brie (2001). Introspecting phenomenal states. *Philosophy and Phenomenological Research*, *63,* 305–328.

Goldman, Alvin I. (2006). *Simulating minds*. Oxford: Oxford.

Gopnik, Alison (1993a). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16,* 1–14.

Gopnik, Alison (1993b). Psychopsychology. *Consciousness and Cognition, 2,* 264–280.

Gordon, Robert M. (2007). Ascent routines for propositional attitudes. *Synthese, 159,* 151–165.

Heil, John (1988). Privileged access. *Mind, 97,* 238–251.

Helmholtz, H. (1856/1909/1962). *Helmholtz's treatise on physiological optics,* ed. J.P.C. Southall. New York: Dover.

Hill, Christopher S. (2009). *Consciousness*. Cambridge: Cambridge.

Hill, Christopher S. (2011). How to study introspection. *Journal of Consciousness Studies, 18* (no. 1), 21-43.

Hohwy, Jakob (forthcoming). Phenomenal variability and introspective reliability. *Mind and Language.*

Horgan, Terence, and Uriah Kriegel (2007). Phenomenal epistemology: What is consciousness that we may know it so well? *Philosophical Issues, 17,* 123–144.

Hurlburt, Russell T., and Eric Schwitzgebel (2011). Presuppositions and background assumptions. *Journal of Consciousness Studies, 18* (no. 1), 206-233.

Hurley, Susan (1998). *Consciousness in action.* Cambridge, MA: Harvard.

Kant, Immanuel (1781/1787/1997) *Critique of pure reason,* Paul Guyer and Allen W. Wood (ed. and trans.). Cambridge: Cambridge.

Locke, John (1690/1975). *An essay concerning human understanding.* Oxford: Oxford.

Lycan, William G. (1996). *Consciousness and experience.* Cambridge, MA: MIT.

McGeer, Victoria (2008). The moral development of first-person authority. *European Journal of Philosophy, 16,* 81–108.

McGeer, Victoria, and Philip Pettit (2002). The self-regulating mind. In *Language and Communication*, 22: 281–299.

Moran, Richard (2001). *Authority and estrangement.* Princeton: Princeton.

Nichols, Shaun, and Stephen P. Stich (2003). *Mindreading.* Oxford: Oxford.

Nisbett, Richard E. and Wilson, Timothy D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Noë, Alva (2004). *Action in perception.* Cambridge, MA: MIT.

Papineau, David (2002). *Thinking about consciousness.* Oxford: Oxford.

Prinz, Jesse J. (2004). The fractionation of introspection. *Journal of Consciousness Studies, 11* (no. 7–8), 40–57.

Prinz, Jesse J. (2007). Mental pointing: Phenomenal knowledge without concepts. *Journal of Consciousness Studies, 14* (no. 9-10), 184-211.

Russell, Bertrand (1912). *The problems of philosophy.* London: Oxford.

Ryle, Gilbert (1949). *The concept of mind.* New York: Barnes and Noble.

Schwitzgebel, Eric (2005). Difference tone training: A demonstration adapted from Titchener's *Experimental Psychology* (1901-1905), vol. I, part 1, 39-46. *Psyche*, 11 (6). [Please check my academic homepage for a link to this web-based demonstration.]

Schwitzgebel, Eric (2010). Introspection. *Stanford Encyclopedia of Philosophy.*

Schwitzgebel, Eric (2011a). *Perplexities of consciousness.* Cambridge, MA: MIT.

Schwitzgebel, Eric (2011b). The philosophical and psychological context of Descriptive

   Experience Sampling. *Journal of Consciousness Studies.*

Shoemaker, Sydney (1996). *The first-person perspective and other essays.* Cambridge:

   Cambridge.

Titchener, E.B. (1901-1905). *Experimental psychology.* New York: Macmillan.

Titchener, E.B. (1910). *A text-book of psychology.* New York: Macmillan.

Tye, Michael (2000). *Consciousness, color, and content*. Cambridge, MA: MIT.

Wittgenstein, Ludwig (1953/1968). *Philosophical investigations*, 3rd ed. G.E.M. Anscombe

   (trans.). New York: Macmillan.

Wundt, Wilhelm (1888). Selbstbeobachtung und innere Wahrnehmung. *Philosophische Studien,*

   4, 292-309.

Wundt, Wilhelm (1896/1897). *Outlines of psychology,* C.H. Judd (trans.). New York: Stechert.