

Self-Understanding in the Face of Divine Thought Insertion

ETHAN RICHARDSON

Modern philosophy began with an instance of selfknowledge. When Descartes published the *cogito*, he asserted that there are some kinds of things we can know about ourselves that just can't be wrong. That claim has been largely disputed since Descartes' day, and different accounts try to clarify what exactly "I think, therefore I am" meant, but the issue of self-knowledge (what we can or do know about our own mental states) remains at the genesis of the discipline in the modern era.

Tyler Burge gives a compelling account of self-knowledge in terms of what he calls self-understanding, as published in the Dewey Lectures of 2007. Burge effectively argues that in cases of moral or critical reasoning, what we know about our own mental states is immune to a certain kind of error he calls "brute error." He maintains that were our self-knowledge in these cases subject to brute error, the epistemic warrants we commonly ascribe to moral and critical reasoning would not apply with the same force. The nature of the warrants tells us something about the nature of self-understanding and, therefore, selfknowledge.

In this paper, I present Burge's account of self-knowledge in the context of David Armstrong's prior influential competing account. I mention a significant counterexample to Burge's claims that many theists will have to reckon with—namely, a specific form of divine

Ethan Richardson is an undergraduate at Brigham Young University studying philosophy and applied math. After graduating in 2028, he plans to pursue graduate school in philosophy. His interests include philosophy of science, logic, and philosophy of religion.

thought insertion (DTI) that I term “revelation incognito.” I motivate this counterexample by presenting psychological findings and appealing to scripture. I explore how DTI might interact with Burge’s framework and draw conclusions from these findings about Burge’s claims regarding brute error and norms of second-order thinking. I finish with an important observation about these norms for anyone who accepts revelation incognito.

Self-Knowledge

One important debate in the self-knowledge literature asks which metaphors we should employ to structure our thinking about self-knowledge, and particularly whether vision is a good metaphor. In his landmark 1968 book *A Materialist Theory of the Mind*, David Armstrong argues that “in introspection, . . . we become aware of current happenings in our own mind” (323). Thus, when we are aware of our mental states, we have a merely empirical priority (front-row seats, as it were) and not a metaphysical or infallible one (325). When one “looks” inside, as the very word “introspection” suggests, one “sees” the contents of the mind, and can come to self-knowledge using an externalist framework for justification.

Tyler Burge takes issue with this purely observational account of self-knowledge; he thinks there are at least some cases in which the nature of our self-knowledge is crucially non-observational. Observation is subject to what he terms “brute error”: “I call a false but well-functioning and warranted propositional attitude, warranted in way *W*, an attitude that incurs *brute error* relative to warrant *W*” (188). Because observation’s success is contingent not only on the proper functioning of the relevant faculties but on other factors such as environmental conditions, even when warranted, representation may fail to be veridical. For example, an expert in illusions might create the image of a dog in front of me. I may be warranted in believing there’s a dog there (for Burge, warrant comes down to something like reliability plus making sense of one’s perspective) and my vision is working properly, but I am wrong about the dog (187–8). This is a classic case of brute error, which Burge thinks cannot happen for a certain kind of metarepresentation that happens in critical and moral reasoning. Burge proposes his own framework for thinking about self-knowledge in critical and moral reasoning that centers on his unique picture of self-understanding in place of an Armstrong-style vision metaphor.

According to Burge, we have certain beliefs at a first-order psychological level, such as that Mozart loved Haydn.¹ We might employ this proposition during normal reasoning, as follows:²

- (1) Mozart loved Haydn.
- (2) If Mozart loved Haydn, then Haydn lived before or during Mozart's lifetime.
- (C) Haydn lived before or during Mozart's lifetime.

According to Burge, critical reasoning (at a second-order or meta-psychological level) takes the beliefs of premises (1) and (2) and represents them as *beliefs*. Thus, during critical reasoning, one might represent (1) as (1*), as in the following example:

- (1*) I believe that Mozart loved Haydn.
- (2) I have no evidence to believe that Mozart loved Haydn.
- (C) I should stop believing that Mozart loved Haydn.

Turning (1) into (1*) involves three simultaneous processes: betokening, indication, and (self)attribution. These correspond to certain parts of the proposition. The proposition I believe that Mozart loved Haydn can be broken into 3 parts. That Mozart loved Haydn is the *content*, belief is the *mode*, and I is the *subject*.³ In critical reasoning, the content is betokened, the mode is indicated, and the two together (the full predicate) are attributed to the subject. Attribution is not a second step after the first two. It all happens together.

The reason Burge thinks we cannot have brute error in this kind of meta-representation is reflected in the nature of betokening representation.

¹ Except in the formalized arguments, I follow Burge's usage of underlining in his Dewey Lectures to indicate mental states. This notation replaces what would probably otherwise be in quotes.

² See Burge's third Dewey lecture, "Self-Understanding", for this account in his own words (I feel compelled to add, at your own peril). I especially focus here on the content he delivers in section V.

³ Burge doesn't explicitly identify the "subject" of the thought (partially, I think, because he does not think that thoughts can be misattributed in this form of representation), but I have added it for the sake of clarity and because it is the dimension of the mental state that I focus on in my subsequent arguments.

The word “betokening” emphasizes that the representational faculty is not representing the thought separately from the thought itself; the faculty reuses the thought token, *de re*, as Burge asserts, in the second-order representation. It’s the difference between copying a printed sentence in the copy machine and adding any additional needed information to the copy versus taking the original sheet and writing on it.⁴ In this way, the betokening cannot get wrong what the belief is, because the belief is *constitutive* of the second-order proposition. This resembles cogito-like cases he examines in which the proposition is guaranteed to be true because it refers to a part of *itself*. In Burge’s language, the critical reasoning and the first-order reasoning are from the same “point of view” (144).

An analogous process involves sight.⁵ At the first-order level, an image is presented to consciousness. One can form beliefs about an “outer” world inductively on the basis of this image, but this is not the kind of belief formation we are concerned with. Instead, we focus on beliefs about the *appearance*. One recognizes the image that is presented to consciousness, then forms in the mind the metapsychological proposition “I see [*image*],” in which [*image*] is the very image presented to consciousness. In doing this, one makes a second-order proposition *using the very first-order* mental state. The proposition is second-order because it involves indication and attribution, and the content is reused (betokened). I is the subject to which the mental image is attributed, see is the psychological mode indicated in the meta-representation, and the image itself is betokened. This is a tricky example because we typically do not syntactically accommodate mental images into propositions, but I hope the analogous process is helpful to illuminate the way Burge is thinking about self-understanding in the more standard critical and moral reasoning cases.

Betokening, indication, and attribution do not exhaust the representational capacities employed in critical reasoning. Other powers are required. For instance, in critical reasoning, one must represent reasons as *reasons in a network*, with relations to other reasons. These important extensions of Burge’s theory are tangential to the main thrust of his epistemic argument, however. Really, he just wants to show that any error in representing first-order reasoning must be a malfunction of critical reasoning and not brute error, because the first-order reasoning is itself contained in the second-order representation.

⁴ If you know C++ or a similar language, it’s the difference between passing an argument into a function normally versus by reference.

⁵ Burge only briefly mentions this example, but I think it is valuable to clarify his earlier theory.

Divine Thought Insertion

Betokening in self-understanding is thus immune to brute error (i.e., if critical reasoning is working right, we can't get the content wrong), but there may be some cases that stretch Burge's account when it comes to attribution. Under normal circumstances, it will be quite all right and warranted to just attribute all mental phenomena to the self. The main difficulty (for theists, at least) arises when one considers God's potential to insert thoughts into the human mind. Psychologists and philosophers have studied thought insertion before, but the phenomenon is typically understood as a symptom of schizophrenia and is therefore analyzed as a pathology. Burge himself acknowledges this literature as a potential objection to his account of the warrant for attribution in self-understanding, but he rejects it on the grounds we just mentioned, saying, "Such cases do not threaten the account because they involve pathology in the preservative capacities that ground entitlements to self-attribution. The schizophrenics do not rely purely on betokening understanding. . . . There is some malfunction" (225). Burge's argument about the nature of self-understanding relies on an external account of the relevant warrants, an account which stipulates that the relevant faculties be functioning properly. Since thought insertion is a pathology, it fails to be a warranted form of attribution and therefore does not threaten Burge's claim that no *warranted* self-understanding in critical reasoning or moral reasoning is subject to brute error.

Thought insertion may derive from other sources, however. Psychiatrists Simon Dein and Christopher Cook examined perceived thought insertion in religious experience in a 2014 study of congregants at an Anglican Church in East London. They discovered that DTI, as opposed to audible voices, was *the paradigm case* of divine communication among the congregants they interviewed (102). They also explicitly address the epistemic question of attribution:

The degree of certainty pertaining to the divine origin of these experiences varied. In some instances, recipients were doubtful that the communication came from God. In other cases, the divine origin, at least to them, seemed blatantly clear. How does a recipient of such experiences know that the communication is from God, rather than from their own thoughts or from other agencies, such as those of Satan? Subjects in this study described a number of ways for discriminating between God's voice and their own thoughts. (11)

These participants reported using phenomenological giveaways, recourse to institutional authorities, and corroboration with scripture (11–12). As the passage suggests, such methods were helpful in some cases, but there appear to be cases

in which the origin of certain thoughts remained dubious—in other words, it was impossible to attribute the thought conclusively. Thus, the findings illuminate a real concern for certainty in thought attribution, a case that cannot be reduced to the pathology of schizophrenia. It is true that most of the cases in the paper do not involve the kind of moral or critical reasoning Burge is talking about, but I think it nonetheless provides good evidence that, from the point of view of a significant number of modern-day theists, God is in the business of thought insertion. This DTI may or may not always be discernible as such to the human psyche. Paradigm cases of DTI which are not discernible as such (a kind I have termed “revelation incognito”) will obviously not be the kind that is reported in a psychiatric study. They do seem represented in accepted Christian and Jewish scripture, however, as surveyed below (passages taken from the NRSV):

Reference	Passage	Gloss
Exodus 10:20	“But the Lord hardened Pharaoh’s heart, and he would not let the Israelites go.”	God inserted a negative attitude towards letting the Israelites go OR God inserted some faulty moral reasoning into Pharaoh’s mind.
Ezra 7:27	“Blessed be the Lord, the God of our ancestors, who put such a thing as this into the heart of the king to glorify the house of the Lord in Jerusalem.”	God inserted an idea and a favorable attitude towards the idea into the mind of Artaxerxes.
2 Thessalonians 2:11	“For this reason God sends them a powerful delusion, leading them to believe what is false.”	God inserts a false belief or God inserts some faulty critical reasoning into the minds of the wicked.

Some of these passages are controversial in the history of Christian biblical interpretation, but I think they serve at least as *prima facie* evidence for revelation incognito. Presumably, neither Pharaoh, nor Artaxerxes, nor the wicked people described in Thessalonians attributed the attitudes and thoughts in their minds caused by God to God. These instances can therefore be taken as instances of revelation incognito. It is also important to note that God appears to be interfering not only in first-order thinking. God may be directly influencing critical or even moral reasoning, as the glosses on the first and third passages demonstrate.

Back to Burge

Given that revelation incognito remains a real possibility for many Christians, how should we evaluate Burge's claims about moral and critical reasoning? If Burge is correct that well-functioning moral and critical reasoning always attribute first-order thinking to the reasoner, then it may be that in cases of revelation incognito, the representational faculties in moral and critical reasoning just always misattribute. Does this mean that the attributive faculty is malfunctioning, or incompetent? I argue that, intuitively, this is not the case. There is no pathology involved in attributing the thought to oneself; it is merely the result of the attributive faculty operating normally, under the unusual circumstance that God inserts a thought and wants (for whatever reason) to keep God's influence hidden. It is rather like a reliable Toyota that one tries to start at the top of Mt. Everest; can we blame the car for not starting given the unusual conditions? I think we should take revelation incognito to be an analogously unusual circumstance for the normally reliable attribution faculty.

It is tempting to think that DTI is always accompanied by a clear phenomenological giveaway. This is what participants in Dein and Cook's study seem most often to point to when pressed to justify their attributions to the divine—this is what makes DTI a paradigm religious experience. However, the study is also clear that confidence in attribution to the divine varies significantly, which suggests that DTI works on a phenomenological spectrum from obvious to indiscernible. It seems clear that if revelation (fully) incognito were a real phenomenon, it would not be reported as such. Just because there is non-contradictory theoretical space open here for revelation incognito does not entail there are actual instances, but the biblical evidence and the commentary in the study provide some reason to think it happens, and they seem to be evidence enough to contradict the generalized phenomenological giveaway claim.

Perhaps Burge's theory of attribution is only trying to say that thoughts that occur in the mind are attributed to that mind by virtue of their locus of occurrence, and they do not make any sort of causal claim about the agency behind the thought. One might point out that Burge's groundwork of ego-centric indexes suggests that phenomena of experience, mental or otherwise, are indexed to the ego-center in a way fundamentally noncommittal about mechanisms or causal agents. Just as one indexes a perceived block of cheese relative to the self physically, so too one indexes thoughts relative to the self mentally. On this view, to be one's own thought is simply to occur within one's mind, regardless of the cause of the thought. I think this must not be what Burge means because it is immediately at odds with our own intuitions about thought

attribution. In asserting “I think that p ,” I typically mean something that includes both a commitment to the thought’s mental locus (my own mind) *and* something about my own causal involvement, something like “I myself am performing the act of thinking that p ,” not merely “ p is occurring within my mental sphere.”

Norms

If DTI threatens the veridicality of self-attribution in critical and moral reasoning even when these faculties are functioning correctly, then we must accept that, at least on a relevant Christian/Jewish picture, the representational faculties involved are subject to brute error. We may not get the content of the thought wrong, but we can indeed misattribute the thought, given that God sometimes inserts revelation incognito.

What does this mean for Burge’s claims about the responsibility this kind of immunity to brute error entails? Burge emphasizes that moral norms rely on the fact that certain psychological points of view have “buck-stopping status.” That is, we can trace full responsibility back to the self that caused some act (and *stop the buck there*) if and only if that self had a well-functioning moral reasoning capacity that could evaluate the psychological antecedent of the act in question (176). Burge’s entire account of moral responsibility rests, then, on his account of self-understanding. Not only is the bare representational capacity important to the norm—so is its immunity to brute error. Burge writes:

If one can understand a psychological element only in a way that is liable to brute error, that element cannot be a basic starting point for assessing the individual’s contribution to full unmitigated accountability under the norms of critical reason and morality. Being subject to brute error means that in reviewing one’s psychological state or event, one is not fully accountable for getting that state or event right. (189–90)

If Burge-style second-order representation is not immune to brute error after all, this has big downstream consequences for the norms. More explicitly, if Burge’s analysis of the norms is right, but brute error is introduced, then there is a fundamental mismatch between what the norms assume about our self-knowledge and what is actually the case. One tempting route for anyone who has accepted my arguments is to reject Burge’s statement here about brute error and claim that the norms still fully apply in spite of brute error. There must have been something wrong with Burge’s analysis of the norms in question.

Let's examine this hypothesis using the first scriptural passage noted above—the hardening of Pharaoh's heart. As noted in the gloss, the scripture accommodates (within Burge's framework) two possibilities for the DTI: either God inserted (1) a first-order negative attitude towards letting the Israelites go (something like not let the Israelites go)⁶ or (2) a second-order faulty moral evaluation of his first order attitude. In the first case, Pharaoh still has a functioning moral reasoning faculty, which will represent the first-order attitude incorrectly as "I don't want to let the Israelites go." This correctly represents the *content* and *mode* of Pharaoh's first-order thought, but it has incorrectly assumed that Pharaoh himself is the source. Since Pharaoh's moral reasoning faculty is functioning correctly, we would expect it to engage in some reasoning like what follows:

- (1) I don't want to let the Israelites go.
- (2) If I don't let the Israelites go, bad consequences will ensue.
- (C) I should ignore (1).

Pharaoh's thought (1) may be incorrectly attributed, but because he is still capable of accurate moral evaluation, he is clearly still responsible for his action. He should have known better than to follow such a thought, even if that thought was externally sourced. This is similar to the way we would treat a case of intrusive thoughts, given their involuntary nature (as they are typically understood). Interestingly, this may indicate that the source or subject of first-order thought does not matter for subsequent responsibility.

In the second case, Pharaoh's negative attitude not let the Israelites go is indeed his own, but in the course of moral reasoning it gets misevaluated, in something like what follows:

- (1) I don't want to let the Israelites go.
- (2) It is right not to let the Israelites go.
- (C) I should act on (1).

On this picture, God has inserted (2) to reinforce Pharaoh's disposition not to let the Israelites go, but (2) is false, so Pharaoh has performed a

⁶ Representations of first-order mental states can seem primitive and not quite well-formed. Note, however, that I don't want to let the Israelites go is already a second-order assertion, presumably borrowing its content from not let the Israelites go.

bad moral evaluation. In this case, there is a good argument to be made that Pharaoh's moral reasoning is being interfered with. Now, we can't blame Pharaoh for his actions in the same way, but Burge is okay with this because he thinks the norms depend on a *well-functioning* moral reasoning capacity and in this case that condition has failed.

The cases reveal some important limits to the extent of revelation incognito's effect on Burge's story about responsibility. If DTI occurs at the second-order level, interference in the reasoning faculty collapses moral responsibility, but this is just like any case of moral reasoning failure. Burge's claims about the importance of immunity to brute error in moral reasoning remain unthreatened, because it wasn't functioning correctly. The first-order case is more interesting. It appears that Pharaoh is still responsible even though he misattributed not let the Israelites go to himself. But is the nature of the responsibility the same? Is Pharaoh responsible in exactly the same way when God inserts the attitude as when the attitude is truly his own desire? Intuitively, I don't think so. It seems Pharaoh's responsibility is mitigated at least somewhat by the fact that the first-order attitude came from God and was not his own.

So, instead of rejecting Burge's claims about brute error's role in moral responsibility, I accept them. The norms are correctly analyzed as precluding the possibility of brute error, and therefore it is also the case that they fundamentally mismatch our self-understanding. Given that we can be wrong in attributing thinking to ourselves, we should be wary to apply full, unmitigated responsibility to any act for which second-order reasoning about a psychological antecedent is a crucial factor. This move allows us to have some clemency for Pharaoh. It validates to an extent the Israelites' praising God instead of Artaxerxes for the important good acts he performed. Revelation incognito reveals that, for Abrahamic theists, our commitment to unmitigated moral responsibility may rest on a shaky foundation.

Conclusion

Descartes was sure that "I think." For some theists, however, sometimes God does the thinking. Our second-order psychological representation faculties are ill-equipped to deal with this kind of intrusion, leading in certain cases to misattribution of these thoughts to oneself. When such cases are psychological antecedents to actions with moral weight, I believe moral responsibility is mitigated. Throughout, I have adopted a Burgean framework for self-knowledge and self-understanding and have shown how certain claims he makes regarding brute error fail. I also borrowed his analysis of moral norms to demonstrate the mitigation claim.

Works Cited

- Armstrong, D. M. "Introspection." In *Self-Knowledge*, Quassim Cassam (ed.), Oxford UP, 1994, pp. 109–17.
- Burge, Tyler. *Self and Self-Understanding*. Oxford UP, 2010.
- Dein, Simon, and Christopher C. H. Cook. "God Put a Thought into My Mind: The Charismatic Christian Experience of Receiving Communications from God." *Mental Health, Religion & Culture*, vol. 18, no. 2, 2015, pp. 97–113.

