

# Externalist Thought Experiments and Directions of Fit

*Casey Woodling*

*Coastal Carolina University*

## *Abstract*

The classic thought experiments for Content Externalism have been motivated by consideration of intentional states with a mind-to-world direction of fit. In this paper, I argue that when these experiments are run on intentional states with a world-to-mind direction of fit, the thought experiments actually support Content Internalism. Because of this, I argue that the classic thought experiments alone cannot properly motivate Content Externalism. I do not show that Content Externalism is false in this paper, just that it cannot be motivated by the classic thought experiments alone. I discuss various externalist responses to the argument I raise and show that they all fail.

*Keywords:* directions of fit, content externalism, content internalism, thought experiment, Twin Earth, Burge

Content Externalism holds that the content of intentional states is not determined solely by the intrinsic or non-relational properties of the subjects who have those states.<sup>1</sup> In short, intentional content fails to be completely determined by the ways a subject is independent of the environment in which he or she is embedded.<sup>2</sup>

<sup>1</sup> Brown (2004) rightly notes that Content Externalism breaks down into different varieties. In this paper, I mainly focus on the two versions that are motivated by the classic thought experiments, though I do discuss other versions of Content Externalism in the paper, namely Fodor's version of Content Externalism related to his Conceptual Atomism. One of the main versions of Content Externalism I focus on is Natural Kind Externalism, which is the idea that intentional states about natural kinds involve natural kind concepts whose individuation depends on factors outside of the subjects who have or grasp these concepts. The other version that I focus on is Social Externalism, which is motivated by Tyler Burge's work. This view is more encompassing than Natural Kind Externalism and holds that the concepts that structure our thoughts depend on our relationship to others and the wider community in which we are embedded. Thus, our intentional contents do not depend on intrinsic properties alone.

<sup>2</sup> This debate has also been put in terms of wide and narrow content. Narrow content is content that is completely determined by a subject's intrinsic properties while wide content is content that is not completely determined by a subject's intrinsic properties but depends

Some philosophers are convinced of the truth of this doctrine because of the classic externalist thought experiments, such as Tyler Burge's arthritis case and the Twin Earth thought experiment. Here is Paul Boghossian on the powerful role these classic thought experiments have had in motivating Content Externalism.

[P]hilosophers who embrace externalism don't do so because they regard it as a self-evident truth. They embrace it, rather, because their intuitive responses to a certain kind of thought experiment—Putnamian Twin Earth fantasies—appear to leave them little choice.<sup>3</sup>

The classic thought experiments ask us to consider cases that involve beliefs, which have a mind-to-world direction of fit.<sup>4</sup> In this paper, I show that when we consider intentional states with a distinct direction of fit, such as desires, the thought experiments actually motivate Content Internalism and not Content Externalism.<sup>5</sup> Looking at intentional states with world-to-mind direction of fit, then,

at some level on the extrinsic or relational properties of the subject. There are dual-factor theories of content, which hold a role for both types of content. Ned Block (1986, 1987) has been very influential in this regard. It is thought that such approaches could do justice to both Content Externalism and Content Internalism by allowing for both sorts of content. Narrow content is needed for psychological explanation and to explain the rationality of subjects while wide content is needed to respect the role that the environment plays in determining content. I myself endorse the view that intentional content is narrow and semantic content is wide. That is, I endorse a view that combines Content Internalism and Semantic Externalism. However, there is no direct argument for this view in the paper, nor is the truth of the view presupposed anywhere.

<sup>3</sup> Boghossian 1997: 163.

<sup>4</sup> Though she did not use the expression 'direction of fit', the conceptual distinction is usually credited to Anscombe 1957. The distinction and terminology also appear in Austin 1953.

<sup>5</sup> The two types of direction of fit I shall consider are mind-to-world direction of fit and world-to-mind direction of fit. These categories are not exhaustive. As John Searle (2004) notes, some intentional states have a null direction of fit. For the purposes of this paper, it is important to understand that states with mind-to-world direction of fit earn their name because the aim of these mental state is for the mind to fit the world. Beliefs are intentional states that clearly have a mind-to-world direction of fit. States with a world-to-mind direction of fit earn their name because the aim of these mental states is for the world to fit the mind. Desires are clear examples of this type of direction of fit. For what it is worth, Searle 2004 offers a brief diagnosis of the classic thought experiments involving beliefs. To the Twin Earth argument, he argues in favor of an internalist intuition that the satisfaction conditions for intentional states are set from the subject's point of view. I think this approach is largely correct, though he suggests that the satisfaction conditions for intentional states about what is called 'water' is largely consistent across the respective populations, a point with which I am not in complete agreement. I would want to allow for more diversity in terms of the intentional content across individual subjects. His diagnosis of Burge's thought experiment is more or less that the relevant intentional content of the two individuals is the same. The only difference is that in one case the patient's use of 'arthritis' diverges from the community norms. Searle says this divergence is fine, but not enough of a reason for thinking that the respective intentional contents are distinct. I am largely in agreement with this diagnosis of the original thought experiment, though I think that in the end more needs to be said to address Burge's Social Externalism properly as it motivated by other reasons, namely that Social Externalism is true if our thoughts connect to an objective reality.

shows that Content Externalism cannot be motivated by the classic thought experiments alone. Though there are many interesting issues in the vicinity, my focus in the paper is rather narrow: to show that the classic externalist thought experiments do not support Content Externalism on their own. This conclusion should be acceptable to both those who see thought experiments as viable tools in philosophical argumentation as well as those who are more skeptical. For the non-skeptic, I present evidence that the classic thought experiments support Content Internalism when run on states with a world-to-mind direction of fit, so the externalist intuition does not hold over various directions of fit. Those who are more skeptical of thought experiments in philosophy can read the conclusion as providing more grist for the mill; there is genuine debate about which intuitive response to the classic externalist thought experiments is best, so we must conclude that the thought experiments alone cannot motivate either Content Externalism or Content Internalism. Either way, the classic thought experiments alone do not motivate Content Externalism.<sup>6</sup>

### 1. The Classic Externalist Thought Experiments

Here is the textbook version, from *The Stanford Encyclopedia of Philosophy*, of how Twin Earth can be extended from Semantic Externalism (a view about linguistic meaning) to Content Externalism (a view about the intentional content of thoughts).

Although this thought experiment was designed to establish semantic externalism, it can be extended to mental contents as well (see McGinn 1977). Thus, consider an individual on Earth who sincerely utters ‘water quenches thirst’ before 1750. Such an individual would be expressing his belief that water quenches thirst, a belief that is true if and only if H<sub>2</sub>O quenches thirst. The externalist then asks us to consider a physically identical counterpart of this individual on Twin Earth. Being a resident on Twin Earth, this counterpart has only encountered twin-water, and has never encountered samples of water or heard about water from other people. According to the externalist, our intuition tells us that this individual on Twin Earth does not believe that water quenches thirst. When he utters ‘water quenches

<sup>6</sup> Those who are looking for an argument for the falsity of Content Externalism will not find one in this paper. Some of the most common arguments of this form appeal to epistemic notions such as privileged access, self-knowledge and first-person authority, holding that Content Externalism is not compatible with agents having these epistemic properties in the correct ways. I also endorse a version of this argument. In a nutshell, the problem is that some forms of Content Externalism face a dilemma on the assumption that the concept associated with a word is what expresses the meaning of that word. The dilemma involves different types of concepts, mental particulars and abstracta. Either the content externalist says that the concepts that constitute intentional content are mental particulars or abstracta. If it is the first option, then the content externalist has to reject the distinction between communal and idiolectic meaning. If the latter option, then the content externalist has to reject the idea that subjects have privileged access to content, since no individual has privileged access to abstracta. A final option for the content externalist is to hold that concepts are neither particulars nor abstracta but reducible to abilities. This is not a live option, though, for Social Externalism (Burge’s form of externalism) or Natural Kind Externalism (externalism motivated by Twin Earth cases) as these views are not consistent with this view of concepts. Not all forms of Content Externalism face the dilemma, but the two major views discussed in this paper do.

thirst', he is instead expressing the belief that twin-water quenches thirst, a belief with different truth-conditions. In short, these two individuals have different beliefs despite being intrinsically identical (ignoring the fact that the human body is about 60% water). It follows that some beliefs do not supervene on intrinsic facts, and therefore that externalism is true.<sup>7</sup>

The twins share all the same intrinsic properties, though their respective intentional states differ in terms of their truth conditions. One twin's belief that water is vital to human life is true if and only if H<sub>2</sub>O is vital to human life, and the other twin's belief is true if and only if XYZ is vital to human life. The difference in the extension of 'water' at each world makes for distinct truth conditions that in turn make the content of their beliefs different. This difference in intentional content cannot be captured by the intrinsic properties of the twins for they share all the same intrinsic properties, so intentional content is not determined merely by a subject's intrinsic properties. Content Internalism is false.

At a minimum, our reflection on the twins and the different substances at their respective worlds shows us that the nature of the environment has a decisive impact on determining intentional content according to the most popular interpretation of the thought experiment. The most basic Twin Earth intuition appears to be: the difference in chemical structure of the watery stuff in each environment makes for a difference in the content of the twins' beliefs about the watery stuff in each of their respective environments. In part, we come to this intuition by way of the set-up of the thought experiment. It is built into the thought experiment that intentional content is determined by conditions of satisfaction (which includes truth conditions). I take the assumption that intentional content is determined by satisfaction conditions to be a harmless assumption in the context of the thought experiment.<sup>8</sup> The conditions of satisfaction of intentional states are the state of affairs in the world that would satisfy the intentional state. In the case of belief, this is the state of affairs that would make some belief true. Because desires are not true or false, the state of affairs that satisfies some desire constitutes the desire's conditions of satisfaction.

Through reflecting on the classic Twin Earth thought experiment, we see that the truth conditions or satisfaction conditions of intentional states can be determined in a rather straightforward way: we note the terms in the intentional state reports, note their respective extensions, and then determine the satisfaction conditions of the intentional states of the respective reports. The intuition seems to be a powerful one.

Tyler Burge offers the other classic externalist thought experiment in which he describes two different patients who both assert, "I have arthritis in my thigh." The patients are embedded in environments where 'arthritis' has distinct linguistic or conventional meanings. In the first patient's environment, it means a rheumatoid ailment exclusively of the joints, and in the second patient's environment it means a rheumatoid ailment of either the joints or the muscles. Because of this difference in linguistic meaning, there is a difference in the truth conditions of the

<sup>7</sup> Lau and Deutsch 2014.

<sup>8</sup> I am not saying here that the content of *all* intentional states is fixed by conditions of satisfaction. See Crane 2013 for resistance to the idea that all intentional content is determined by conditions of satisfaction.

respective beliefs, and therefore a difference in their intentional content. The individuals share all the same intrinsic properties, so intentional content fails to supervene on intrinsic properties. The dominant intuition about this case is that Burge's externalist read of the thought experiment is indeed correct.<sup>9</sup>

There is a minority report. In the context of describing a version of the thought experiment where Twin-Oscar comes to Earth, Noam Chomsky registers the contrary intuition.<sup>10</sup>

Turning to 'content of belief', if Twin-Oscar continues to ask for what comes from the faucet to quench his thirst, calling it 'water', has he changed his beliefs about water—irrationally, since he has no evidence for such a change? Or is he behaving rationally, keeping his original beliefs about water, which allow for the stuff on Earth to be water (in Twin-English) in the first place? If the latter, then beliefs about water are shared on Earth and Twin-Earth, just as on either planet, beliefs may differ about the very same substance.<sup>11</sup>

Chomsky's intuition is that the beliefs of an individual switched between Earth and Twin Earth would not change based on environmental changes—a thought shared by other internalists. Instead of attempting to sort out these conflicting intuitions, I shall simply grant the externalist intuition about intentional

<sup>9</sup> Burge's defense of Content Externalism is sophisticated and subtle. Burge has been clear that his various externalist thought experiments are deeper than the idea that the semantic content of the ascription of an intentional state always faithfully determines the intentional content of that state. See Burge 2003, 2006 for discussion of this point. He allows room for malapropisms and other instances where subjects misspeak, where, in other words, the standard semantic content of a term or phrase does not properly express the subject's intentional content. When Yogi Berra said, for example, "Texas has a lot of electrical votes," Burge's view is not that we attribute to him the concept ELECTRICAL as a component of his intentional content. In short, the semantic content of ascriptions does not always completely fix a state's intentional content according to Burge. Any plausible version of Content Externalism will have to accept this fact, and Burge's of course does.

<sup>10</sup> Some of Chomsky's philosophical and linguistic commitments, namely his focus on I-language as opposed to E-language, may indeed skew his intuitions here. This is likely the case, given that Chomsky (1986, 1995) holds that I-languages should be seen as the proper object of linguistic study and these objects can be studied without regard to the environment in which subjects are embedded. This point opens up a more general worry about thought experiments: that they merely reveal prior commitments of theorists and do not provide evidence for theoretical commitments. If this is so, then we have a general reason to doubt that the externalist thought experiments motivate Content Externalism, as they would not provide evidence *for* a view but evidence that a view is antecedently held. I do not want to endorse this general skepticism. I bring up the point merely to say that the externalist cannot reject Chomsky's intuitions on these grounds while maintaining the conclusion that Content Externalism is properly supported by the classic thought experiments. Related to this point, Machery (2012) points to data that suggests that one's field (and therefore theoretical focus) can bias one's intuitions about reference. It appears that philosophers of language, for example, have more Kripkean intuitions than do sociolinguists who typically have more descriptivist intuitions. Machery's favored explanation of this data is that one's theoretical commitments bias one's intuitions. I thank an anonymous referee for helping me bring out this important point.

<sup>11</sup> Chomsky 2000: 149. For what it is worth, Chomsky also expresses a general skepticism about the viability of using such thought experiments to tell us anything interesting about language and the mind.

states with a mind-to-world direction of fit for now and ask whether or not the externalist intuition can be sustained over intentional states with distinct directions of fit, for it must be so sustained if the thought experiments are to properly motivate Content Externalism. Let us turn now to intentional states with a world-to-mind direction of fit.

## 2. Testing the Classic Externalist Thought Experiments with Desire

In this section I show that the classic thought experiments elicit internalist intuitions when we run them on intentional states with a world-to-mind direction of fit. Consider a variant of the switching case.<sup>12</sup>

I desire a drink of water. According to the content externalist, on Earth my desire to drink water is satisfied if and only if I drink H<sub>2</sub>O, just as my belief that water is vital to human life is true if and only if H<sub>2</sub>O is vital to human life. Now suppose that I am switched without my knowledge to Twin Earth, per the familiar slow-switching scenario. Here I add one detail to the standard switching scenario. I have a bottle filled with water (that is, H<sub>2</sub>O) that gets transported with me when I am switched. Some time passes, the concept TWATER takes hold such that when I assert, 'I want a drink of water', my desire is now satisfied if and only if I drink XYZ according to a content externalist. In an attempt to satisfy my desire I drink from the bottle that travelled with me from Earth. As far as I am concerned, my desire is satisfied by the event of my drinking from this bottle. It certainly seems to me that my desire is satisfied.

It turns out that my desire is not satisfied by the event of my drinking from the water bottle according to the externalist, because the satisfaction conditions of my desire are not met, for my desire is satisfied if and only if I drink XYZ. Imagine my surprise when I learn that I must drink another glass of what appears to me to be identical to stuff that I just drank in order to satisfy my desire.

Based on the original version of the thought experiment (where we note that the differences in the extension of 'water' make for a difference in intentional states truly described by ascriptions with that term), the content externalist analysis is that my desire can be satisfied only by XYZ and nothing else. So, it would seem that this case shows that the externalist idea that the satisfaction conditions of intentional states are fixed by environmental factors is false, because if Content Externalism is true, then once the new Twin Earthian concept TWATER takes hold, all my intentional states about watery stuff will have the concept TWATER as a constituent. It seems clear, though, that my desire for what I call 'water' will be satisfied by H<sub>2</sub>O. After all, if this stuff satisfied my desire for water on Earth, why would it not satisfy my desire for water on Twin Earth? But this is not something that the content externalist who wants to use the thought experiment to motivate Content Externalism can allow for. The case shows that the conditions of satisfaction according to Content Externalism are too restrictive. It is perfectly obvious that H<sub>2</sub>O (and not just XYZ) will satisfy my desire for what I call 'water'.

<sup>12</sup> The switching scenario was first introduced in Burge 1988. Burge (1988) articulates the commonly held externalist idea that it would take a certain period of unspecified time for a switched individual's concepts to switch over and begin to fit the new environment.

The switching variants of Putnam's original Twin Earth case are not part of the original thought experiment. Externalists divide on how to best handle these cases, so it is important to discuss these divisions. There are, broadly speaking, two approaches to the slow switching cases, the Conceptual Addition Interpretation (hereafter CAI) and the Conceptual Replacement Interpretation (hereafter CRI).<sup>13</sup> CRI holds that after the switch, the switcher's concepts are completely replaced. In other words, if I am switched from Earth to Twin Earth, then my concept WATER, after the requisite period, gets replaced completely by the concept TWATER. On the other hand, CAI holds that both concepts (WATER and TWATER) can be retained; which concept is deployed will depend on the context. It may seem that CRI is the version of externalism that gets the wrong reading of my thought experiment and therefore the externalist can simply accept CAI to avoid the internalist intuition.

This is a plausible move, because I have so far been assuming that content externalists adopt CRI. At this point, we can see that Content Externalism supplemented with CRI cannot be used by someone who wishes to use the classic Twin Earth thought experiments to motivate Content Externalism. But what about Content Externalism supplemented with CAI? If I can acquire the concept TWATER while retaining the concept WATER during my strange trip, then the content externalist can say that my desire for water is satisfied by H<sub>2</sub>O because my desire deploys my concept WATER because the intentional state is in some way tied to my home environment.

It turns out, though, that Content Externalism plus CAI cannot be used by someone who holds that Content Externalism is motivated by the classic Twin Earth thought experiment. To see this point, consider the standard telling of the Twin Earth story, which considers beliefs about watery stuff had by twin individuals. For this scenario to motivate Content Externalism it must be the case that the concept that each individual associates with 'water' not be disjunctive between XYZ and H<sub>2</sub>O, because if it were, the intentional content would clearly not be distinct for each individual. And we need the intentional content of the beliefs to be distinct in order to conclude that intentional content fails to be determined by the intrinsic properties of the subjects. So if the content externalist says that the concept associated with 'water' is disjunctive, then he can say the right thing about my examples, but he cannot use Twin Earth to motivate Content Externalism, since such use of the thought experiment requires that the intentional content be distinct for the twin individuals who have the same intrinsic properties.

The point here is not that the externalist is blocked from ever saying that a single concept can pick out distinct things. The problem for the content externalist in taking this option is related to the use of the classic thought experiments in motivating Content Externalism. In those thought experiments, the externalist cannot say that the relevant concepts are disjunctive because then the thought experiments cannot be used to motivate Content Externalism. A content externalist who adopts this line must say, then, that the thought experiments by themselves do not properly motivate Content Externalism, but that the view can be motivated otherwise, but this is in line with the very conclusion I am arguing for. My point is that Content Externalism cannot be motivated by the classic thought experiments alone.

<sup>13</sup> I use terminology from Parent 2013 here.

To this the externalist who adopts CAI may push back by saying that in the original thought experiment, the question of switching is not raised, so the issue of which concept associated with 'water' gets tokened does not even arise. But this reply misses something crucial about CAI: that the method of determining subjects' intentional content is more difficult than might first appear and involves more context than is available in the original telling of the thought experiment. In the original telling of the thought experiment, we move from the linguistic meaning of the term 'water' directly to the truth conditions of the respective intentional states. There is no consideration of other factors. It is stipulated that the twins' histories are the same, but these are in no way examined to determine the nature of their concepts, and this is precisely what CAI requires, as it embodies a method of content attribution that is subtle, requiring consideration of relevant context. So, in the end, an externalist who adopts CAI ultimately must admit that the original version of the thought experiment cannot alone motivate Content Externalism.

At this point, it is natural for the content externalist to shift to Burge's famous thought experiment and give up on the original version of the Twin Earth one, the idea being that the motivation for Content Externalism may come by way of Burge's thought experiment and the problems just mention can be simply bypassed. To begin to address this move, let us begin by considering a variant of Burge's famous arthritis case.

Suppose that Burge's patient (the patient who, according to Burge, falsely believes he has arthritis in his thigh) also desires relief from pain in his thigh that he reports as being arthritis. He asserts, 'I want the arthritis in my thigh to go away'. As in the original telling, the patient is embedded in an environment where 'arthritis' refers exclusively to ailments of the joints. Also following the original version, we imagine a second patient, one who is an internal duplicate of the first but embedded in an environment where 'arthritis' refers to ailments of the muscles and the joints. The individuals are the same internally. Like the first patient, the second patient asserts, 'I want the arthritis in my thigh to go away'.

What are the externalist intuitions about the conditions of satisfaction of the respective desires? Perhaps the most natural thought is that the desire of the first patient is satisfied just in case the arthritic pain in the patient's thigh stops, and the same could be said for the second patient if we use their words as a guide for what they desire. However, things are more complicated, as 'arthritis' has different meanings in the different environments. In the version with states with a mind-to-world direction of fit, it turns out that, regarding the respective beliefs about what is called 'arthritis', the first patient's belief is false while the second patient's belief is true. So, analogously—stating the satisfaction conditions from our language community—in the present version with states with a world-to-mind direction of fit, the first patient's desire for relief is satisfied if and only if the arthritis in his thigh ceases, while the second patient's desire is satisfied if and only if the tharthritis in his thigh ceases. To make the distinction here vivid, let us suppose that the respective arthritic pain ceases in both cases. On the content externalist analysis, the first patient's desire is not satisfied while the second patient's is. They are the same in terms of intrinsic properties, and assuming that conditions of satisfaction fix intentional content, the respective contents are distinct. Thus, Content Internalism is false.

Is it plausible to attribute to the first patient a desire with conditions of satisfaction that cannot be satisfied because it is built into the intersubjective concept ARTHRITIS that arthritis cannot occur in the muscles?<sup>14</sup> There may be reasonable grounds for attributing to subjects conditions of satisfaction that cannot in fact be satisfied. If someone has a desire to find a very specific idealized soul mate, then we are warranted in attributing to him or her conditions of satisfaction that cannot be satisfied. If a child desires that Santa bring him a new pony, then we are likewise warranted. However, in the case of the first patient, it seems that he has a desire that can be satisfied, so there is no good reason for attributing to him an intentional state with conditions of satisfaction that cannot be satisfied. The internalist intuition, then, is that his desire is satisfied just in case the pain in his thigh that feels to him like arthritis ceases. This can be satisfied or unsatisfied, depending on how the world turns out. In fact, the internalist idea is that the intentional state of the second patient has the very same conditions of satisfaction.<sup>15</sup> The internalist reading attributes conditions of satisfaction that respect the subjects' respective interests on the world and also capture the proper level of detail in their thoughts about their pain. What's crucial to them is not the correct medical classification of their pain, but the pain itself and their prior experiences with such pain. The main reason for preferring the internalist interpretation to the externalist one is that it attributes conditions of satisfaction for the patient's desire about his pain that can be either satisfied or not satisfied. In some cases, as I noted, it is right to attribute conditions of satisfaction that cannot be satisfied (when subjects have desires about fictional entities, for instance). Surely the patient's desire is not about a fictional entity. It is about a very real pain in his thigh. So, the proper reflective response to the above variant of Burge's original thought experiment is an internalist one. Therefore, the version of Burge's thought experiment that involves desires does not support Content Externalism.<sup>16</sup>

The dialectical context is this. The thought experiments most commonly used to motivate Content Externalism elicit internalist intuitions when we consider intentional states with a world-to-mind direction of fit. Of the Twin Earth

<sup>14</sup> When necessary I distinguish between two types of concepts, following Laurence and Margolis 2007, intersubjective concepts that are abstracta and subjective conceptions that are mental particulars.

<sup>15</sup> Note that these intuitions may be arrived at only after a period of reflection. Many of these thought experiments are so unusual and complex that it takes a period of reflection to form one's judgment about them.

<sup>16</sup> It appears that determining the conditions of satisfaction for intentional states with a world-to-mind direction of fit cannot be done without regard to how such states relate to other intentional states in the subject's mental economy. So, it may be that considering intentional states with this direction of fit has ramifications for issues related to Conceptual Holism, Molecularism, and Atomism. This paper is neutral on which of these views is correct. However, a very quick argument for the truth of Conceptual Molecularism (or perhaps something more radical) runs as follows. The conditions of satisfaction for my desire for eating biscuits, say, depend on my beliefs about biscuits. I may have an idiosyncratic idea of what counts as a biscuit for instance. So, if the content of desires is cashed out in terms of their conditions of satisfaction, and those conditions of satisfaction depend for their fixing on other intentional states, then it would seem that at least Conceptual Molecularism would be established. I sketch this argument not to endorse it, but to show that considering intentional states with a world-to-mind direction of fit may be important to do across a range of views in the philosophy of mind. I thank an anonymous reviewer for helping me see this connection.

case, either the content externalist adopts CRI and cannot adopt the plausible read of the thought experiment (that is, my desire is satisfied by the water in my bottle) or the externalist adopts CAI and cannot hold that the original thought experiments motivate Content Externalism, as CAI admits that we need more context than we are given in the original example to determine the intentional content of the respective twins. Of the Burge case, the content externalist must attribute to the first patient a desire with conditions of satisfaction that cannot be satisfied. Sometimes, as I note, this makes sense, such as in cases of desires related to fictional entities. However, in the revised Burge case, on reflection, it is most plausible to attribute conditions of satisfaction that can or cannot be satisfied. So, to be clear, the conclusion at this point is not that Content Externalism is false or that Content Internalism is true; it is that one must appeal to more than the classic externalist thought experiments to support Content Externalism.

### 3. Content Externalist Replies

The content externalist could always accept the conclusion and hold that more than appeal to the thought experiments is needed to properly support Content Externalism. For those content externalists who wish to resist this conciliatory line of thought, there are various strategies for dealing with the variants of the famous thought experiments above. I shall review five replies and argue that all of them fail.

#### *3.1 Limiting the Scope of Content Externalism to Merely Intentional States with a Mind-to-World Direction of Fit*

A content externalist could grant that Content Externalism is true of just states with a mind-to-world direction of fit and not of states with a world-to-mind direction of fit. This concessive move does not work, because it has serious problems when it comes to explaining simple bits of reasoning behind intentional action. On this move we would say that when I am on Twin Earth *unawares*, I desire a drink of water and I believe that twater will come out of the faucet in my kitchen, and this belief and desire pair causes me to go to the kitchen for a drink of water. However, saying this is problematic. What is it that I go into the kitchen for? It does not appear that the externalist has a uniform answer to this question. In short, I would not be able to reason using intentional states with different directions of fit—even though ascriptions of these states would involve the same term in the content-clauses—for the concept expressed (the concepts WATER and TWA-TER) would vary depending on the direction of fit of the intentional state.<sup>17</sup> I need to be reasoning with the same concept in each one of the contents for the contents to properly link up and properly cause my action.<sup>18</sup>

<sup>17</sup> The same problem would arise in the Burgean cases. If the subject desires to rid himself of the arthritis in his thigh and believes that the doctor can help him get rid of it, then for those two intentional states to cause him to make a doctor's appointment, the intentional content expressed by 'arthritis' must be the same for both states.

<sup>18</sup> This problem is similar to the one raised in Boghossian 1992. Boghossian argues that if Content Externalism is true, then subjects cannot detect the validity of their reasoning a priori. So, Content Externalism cannot be squared with the obvious truth that the validity of reasoning should be detectable a priori.

### 3.2 The Kripkean Strategy

Another externalist reply is to hold that the internalist intuitions elicited above rest on conflating, in each case, two distinct desires. In the water bottle case, the H<sub>2</sub>O from the bottle satisfies the subject's desire to quench his thirst, but it does not satisfy his desire for water. In the arthritis case, the patient's desire for the pain in his thigh to cease is satisfied if it ends, but the pain ceasing does not satisfy his desire for the pain from the arthritis in his thigh to stop. I call this the Kripkean strategy because it is similar to the strategy he employs in explaining modal illusion. According to Kripke, one might take oneself, for example, to have imagined water being XYZ, but what they have really imagined is not water but a water-like substance being XYZ. I think that Kripke's strategy applied to modal illusion is implausible, just as I think that this reply is implausible. Let me explain.

So far I have merely granted the externalist intuitions about intentional states with a mind-to-world direction of fit. Responding to this reply helps us to see that these intuitions are mistaken as well. Consider the water bottle case. What is the concept most appropriate to attribute to me when I desire a drink of water on Twin Earth? The descriptive concept CLEAR, COLORLESS, NEARLY ODORLESS AND TASTELESS LIQUID or the natural kind concept WATER which is the concept CLEAR, COLORLESS, NEARLY ODORLESS AND TASTELESS LIQUID AROUND HERE THAT HAS CERTAIN MICROSTRUCTURAL PROPERTIES THAT DETERMINE ITS ESSENCE. Note that the natural kind concept TWATER has a similar description but it is distinct in virtue of the different microstructural properties of twater. The externalist says that it is not the descriptive concept but the natural kind concept that structures my intentional states. Why must we hold such a wooden view of concept attribution, though? Perhaps on different occasions I have states with merely the descriptive concept and perhaps I have states with the natural kind concept. Surely which concept is deployed—the concept WATER, the concept TWATER, or the concept CLEAR, COLORLESS, NEARLY ODORLESS AND TASTELESS LIQUID—depends heavily on the context and exactly how I am thinking about the objects in my environment. But the content externalist who motivates Content Externalism by using the classic thought experiments and holds CRI leaves no room for this sort of context dependence. For those externalist who adopt CRI, in the Twin Earthian environment, the concept TWATER constitutes in part the subject's intentional content. Period, end of story. The point of my variation of the thought experiment is to bring out that sometimes the subject's interest in the world means that the descriptive concept and not the natural kind concept gets tokened. If I want to drink water, it typically does not matter to me whether it is H<sub>2</sub>O or XYZ given that these microstructures give rise to the very same functional and appearance properties. Upon drinking the water in my bottle, I take my desire to be satisfied and it is. Upon being told the whole story about water and twater, I will not revise my belief that my desire is satisfied. The externalist read is that I will admit that I was wrong about whether my desire was satisfied, but I do not think that this is right. It seems open for me to retain my belief that my desire was satisfied given that I got the very thing that I was looking for. In this case, my thinking about the liquid in my environment was not as detailed as the natural kind concept WATER. Surely, we should attribute to subjects concepts that most closely capture how they think of the world. Content Externalism plus CRI does not allow for this, because to all subjects who have intentional states described by

ascriptions involving ‘water’ on Twin Earth, we must attribute the concept TWA-TER (after of course the requisite period of time passes).

Again, those content externalists who endorse CAI have resources to allow for more sophisticated interpretations of intentional content; however, the point for our purposes is that these individuals—due to these very resources—cannot use only the classic thought experiments to motivate their view. They must allow that we do not have enough context in the original thought experiment to determine the nature of the respective intentional contents.

In the Burge case, which concept structures the patient’s thought? The concept ARTHRITIS or the concept PAIN THAT FEELS LIKE ARTHRITIS FEELS? Burge’s view is that it is the former, in part because the patient uses ‘arthritis’ to describe his own intentional state, and this expresses the intersubjective concept ARTHRITIS in this situation. Running the case with desires helps us to better see that it is much more natural and intuitive to think that the concept PAIN THAT FEELS LIKE ARTHRITIS FEELS structures his thought because it allows that his desire is satisfied when the pain goes away. What is crucial to his perspective is that his pain, which feels to him like arthritis, ceases. So, we need not attribute to him the concept ARTHRITIS in this case because he is clearly not thinking of his pain according to the intersubjective concept ARTHRITIS, as this would require that he thinks he has a pain in his thigh that cannot occur in his thigh. As with the case of WATER and TWATER, Content Externalism can be seen to have a much too restrictive view of the concepts that structure the thoughts of subjects. Sometimes we think with natural kind concepts but sometimes we do not. And what makes the differences is not any environmental factors but factors about our interest in the world. Running the thought experiments on desire gives us a new perspective on these thought experiments and helps us to see that Content Externalism supplement with CRI is too rigid and restrictive when it comes to the sort of concepts that we can attribute to subjects.

### 3.3 *Conceptual Holism, Molecularism, and Atomism*

It may seem that the content externalist has other resources to use in responding to the thought experiments involving states with a world-to-mind direction of fit. One thought is that these experiments assume a version of Conceptual Holism or Molecularism and thus do not trouble the conceptual atomist.<sup>19</sup> The defining feature of Conceptual Atomism is that concepts have no internal structure. They

<sup>19</sup> Conceptual Holism is the analog of Meaning Holism. On such a view, the content of any concept depends on its relation to potentially all concepts in an agent’s mental economy. Conceptual Molecularism is a restricted form of Holism. A concept has its content in virtue of the relations it stands in to a restricted range of concepts. Unlike these views, which share the idea that conceptual content is interdependent between concepts, Conceptual Atomism holds that concepts have content solely in virtue of relations they bear to objects in the environment. Fodor is the most prominent defender of this view (see Fodor 1987, 1990 and 1998 for classic discussions and defenses). It would seem that Conceptual Holism and Conceptual Molecularism are perhaps more suited to Content Internalism while Conceptual Atomism is more suited to Content Externalism, given that the former focus primarily (though perhaps not exclusively) on internal relations for fixing intentional content while the latter focuses exclusively on external relations for fixing intentional content. I am not saying that some of these views are incompatible with others, but merely suggesting that Holism and Molecularism seem *prima facie* more suited to Content Internalism while Atomism seems *prima facie* more suited to Content Externalism.

have no components, and get their content from relations that they bear to the environment. Fodor's Asymmetry Dependency Theory has it that concepts have the content they do in virtue of the external factors that typically cause them to be tokened. Of course, this view has historically struggled with explaining misrepresentation. For example, I may think I see a cow in the distance, and deploy the concept COW when there is in fact not a cow in the distance but a dog. To this Fodor holds that the concept COW means cow and not dog because the concept would not be tokened by a person who sees a dog unless it were usually tokened when a person sees cows. Unless my concept COW had been caused by regular interaction with large domesticated animals, then it would not be able to be mistakenly deployed (as it may be on rare occasion) when I see a dog.<sup>20</sup> Obviously, the question of which one of the above views of concepts is correct is outside of the scope of this paper. It is also worth noting that while Conceptual Atomism is taken to be a type of Content Externalism, it is not a form of Content Externalism that is motivated by the classic thought experiments.<sup>21</sup> These thought experiments motivate Natural Kind Externalism and Social Externalism, so it is worth noting that the conceptual atomist-cum-content externalist is not really troubled by anything I say in this paper. My target are those individuals who hold that Natural Kind Externalism or Social Externalism can be motivated by the classic thought experiments alone.

It is worth bringing out, though, that none of my discussion above assumes that concepts are structured entities. I happen to think that they are, but nothing hangs on that above. I do discuss descriptive concepts and thereby discuss concepts that the conceptual atomist would clearly reject. However, I do not beg the question against the externalist by assuming that all concepts are descriptive concepts; I merely examine them alongside other types of concepts and ask the reader to reflect on their plausible deployment in certain situations.

### *3.4 The New Thought Experiments Are Too Complicated*

Another reply is that the versions of the classic thought experiments that I have offered are so strange and complex that they fail to elicit any solid intuitions. Due to their complexities there simply are no natural or intuitive responses. The content externalist who raises this worry would have to bring the point home and say that the original versions of the thought experiments, as well as the switching variants, are too strange and complex to motivate Content Externalism in the first place. These are all very strange situations that admittedly stretch the normal application conditions of our concepts. So, this move by the content externalist, in the end, admits that I am right: something other than the classic thought experiments must be used to properly support Content Externalism.

### *3.5 Additional Worries, Additional Distinctions*

Let me address a final potential worry. One might worry that I am conflating conditions of satisfaction being satisfied based on the subject's interests and the subject believing that conditions of satisfaction are satisfied. There are no doubt cases where subjects believe that their desires are satisfied while they are not. It is crucial to distinguish between an intentional state being "satisfied" because its

<sup>20</sup> See Fodor 1987, 1990.

<sup>21</sup> See Rives 2010.

subject believes it to be and the conditions of satisfaction actually being satisfied. Consider an example where these notions come apart. Suppose I set out to buy my wife a diamond necklace. I go to the jeweler and get a great price on what I believe to be a diamond necklace. Of course, my desire to buy my wife a diamond necklace is satisfied if and only if I buy her a diamond necklace. In this case it is very important to me that the necklace have diamonds and not a superficially similar material. The jeweler it turns out has tricked me: there is not a single diamond in the necklace—just cubic zirconia. So, I believe my desire is satisfied, but it is really not. Perhaps the above water-bottle case is like this, the externalist may argue: I believe my desire is satisfied, but it is really not. Surely my taking my desire to be satisfied is not enough for it to be truly satisfied. The variants of the classic thought experiments I offer do not assume that whether desires are satisfied depends on whether subjects believe that they are. Here is how to tell the difference between the conditions of satisfaction being fixed from the subject's perspective and actually being satisfied and being "satisfied" merely because the subject believes they are. In the case where I buy a cubic zirconia, once it comes to my attention that the necklace contains no diamonds, I will correct my belief that my desire has been satisfied. From my point of view, the desire for a diamond necklace is satisfied if and only if the necklace I buy contains real diamonds. (Unlike the water bottle case, the microstructure here matters very much and is relevant to my interests.) It seems that I will make no such correction when it comes to my attention that the liquid I drank was H<sub>2</sub>O and not XYZ. It seems that I will maintain that my desire for water was satisfied by the H<sub>2</sub>O in my bottle—even when I am given the full information about the distinct chemical composition of the types of watery stuff in my environment. From my point of view, the desire for what I call 'water' will surely be satisfied by H<sub>2</sub>O. Whether my desire for water is satisfied is not determined by whether I believe it to be satisfied. The fact that I can meaningfully examine whether my belief about the satisfaction of my desire is true or false when I become aware of the chemical make-up of the liquid in my bottle shows that there are facts of the matter outside of my believing some way or other which settle the matter about whether the conditions of satisfaction are met. Rather than my desire being satisfied by my belief that it is, it is satisfied because the conditions of satisfaction are met. Even though the conditions of satisfaction are grounded in the subject's perspective, it should be clear that this is not the same thing as saying that the desire is satisfied just in case the subject believes that it is.

The same test can be applied to the arthritis case. Provide the subject in question with the full information about the situation, and then ask whether he would change his belief about whether his desire was satisfied. Suppose that the pain in the subject's thigh ceases; from his perspective, then, his desire becomes satisfied. Now suppose that he also becomes informed that 'arthritis' in his community means a rheumatoid ailment exclusively of the joints. Does he then change his mind about whether or not his desire was satisfied? Surely he will continue to believe that it is satisfied even when informed about his misuse of the term 'arthritis'. He will say that he mistakenly used the term 'arthritis' to describe his desire, but he will not revise his belief that his desire for relief from the arthritis-like pain in his thigh is satisfied. His pain, after all, is gone. His merely believing that his desire is satisfied does not make it satisfied. As in the previous case, the

desire is satisfied because the conditions of satisfaction are fixed from his perspective on the world and these conditions of satisfaction are met when the pain in his thigh—that feels to him like arthritis—ceases.

Though it is not novel, consider a final point to make the project of this paper more acceptable to philosophers who consider themselves externalists. As for the first water-bottle case I discuss, although my desire is not satisfied by XYZ as the content externalist says it is, it can still be true that the meaning of ‘water’ at Twin Earth is distinct from the meaning of ‘water’ at Earth. It may be that the semantic content or linguistic meaning of the intentional state ascription, ‘I desire a drink of water’, differs from place to place. If I am switched unawares between Earth and Twin Earth, I may mean one thing (in the sense of linguistic meaning) when I report my desire on Twin Earth and mean another when I report my desire on Earth, while all the while the intentional content of my desire remains the same in both places. A view such as this would be a combination of Semantic Externalism and Content Internalism and allow that the semantic content of some intentional state ascriptions can fail to properly describe the intentional content of such states. In the first Twin Earth variant, such a view says that I misdescribe my desire as one for water, because ‘water’ as uttered on Twin Earth is understood by the semantic externalist to refer exclusively to XYZ and my desire is clearly satisfied by H<sub>2</sub>O. So, the semantic content of the language of my report is that I desire XYZ. However, this semantic content fails to describe the intentional content of the desire because the desire is clearly satisfied by H<sub>2</sub>O and not merely XYZ.<sup>22</sup>

#### 4. Conclusion

Although the classic externalist thought experiments typically appeal to just beliefs, we should test our intuitions about the classic externalist thought experiments on other intentional states with distinct directions of fit. Although I have focused on desires here, it seems that other types of intentional states with a world-to-mind fit would cause us to have internalist intuitions as well. Consider a modification of the first water-bottle case above. Suppose I hope that there is some water in my backpack. My hope is satisfied just in case there is some water in my backpack. Does the H<sub>2</sub>O in my backpack satisfy my hope? It seems obvious that it does. The content externalist who wants to use the classic thought experiments alone to motivate Content Externalism says that the hope is satisfied only by XYZ if I have been switched to Twin Earth and been embedded in that environment long enough.<sup>23</sup> This is surely the counterintuitive verdict. Our intuitions tell us that the hope for what I call ‘water’ can indeed be satisfied by H<sub>2</sub>O.

We can draw a moral at this point: things go wrong when we fix the satisfaction conditions of intentional states without proper consideration of the subject’s perspective and interests about the world in some context, and we can see this point more readily when we reflect on intentional states with a world-to-mind

<sup>22</sup> See Loar 1988, Ludwig 1996 and Bach 1997 for similar ideas about how semantic content can at times fail to properly capture a subject’s intentional content. My points here should not be understood as being novel, as the distinction between semantic and intentional content has been in the literature for some time. I merely use the distinction to bring clarity to the discussion, as sometimes it is not always kept in mind.

<sup>23</sup> Remember that only the content externalist who endorses CRI can use the classic thought experiments to motivate Content Externalism.

direction of fit. Our focus when we reflect on states with a world-to-mind direction of fit is first on how the subject conceives things as opposed to when we reflect on states with a mind-to-world direction of fit where there is a greater temptation to focus first on how the world is and only secondarily on how the subject conceives of the world or takes it to be.

Running the thought experiments on desires and other states with a world-to-mind direction of fit helps us see that Content Externalism cannot be properly motivated by the thought experiments alone. Let me end by saying that nothing I have said impacts the Twin Earth thought experiment's ability to support Semantic Externalism—the view it was originally designed to support. Perhaps the reported strength of the externalist intuition so often discussed in the literature is a result of running together the thought experiments' ability to support Semantic Externalism with their ability to support Content Externalism. We should be mindful of this, of course, and run the thought experiment separately for each version of externalism and also run it on a variety of intentional states and not just beliefs.<sup>24</sup>

#### References

- Anscombe, E. 1957, *Intention*, Cambridge, MA: Harvard University Press.
- Austin, J. 1953, "How to Talk—Some Simple Ways", *Proceedings of the Aristotelian Society*, 53, 227-46.
- Bach, K. 1997, "Do Belief Reports Report Beliefs?", *Pacific Philosophical Quarterly*, 78, 215-41.
- Block, N. 1986, "Advertisement for a Semantics for Psychology", *Midwest Studies in Philosophy*, 10, 615-78.
- Block, N. 1987, "Functional role and truth conditions", *Proceedings of the Aristotelian Society*, 61, 157-81.
- Boghossian, P. 1992, "Externalism and Inference", *Philosophical Issues*, 2, 11-28.
- Boghossian, P. 1997, "What the Externalist Can Know A Priori," *Proceedings of the Aristotelian Society*, 97, 161-75.
- Brown, J. 2004, *Anti-individualism and Knowledge*, Cambridge, MA: MIT Press.
- Burge, T. 1979, "Individualism and the Mental", *Midwest Studies in Philosophy*, 4, 73-122.
- Burge, T. 1988, "Individualism and Self-Knowledge", *Journal of Philosophy*, 85, 649-63.
- Burge, T. 2003, "Phenomenality and Reference: Reply to Loar", in Hahn, M. and Ramberg, B. (eds.), *Reflections and Replies*, Cambridge, MA: MIT Press, 435-50.
- Burge, T. 2006, "Postscript to 'Individualism and the Mental'", in T. Burge, *Foundations of Mind*. Oxford: Oxford UP, 151-81.
- Chomsky, N. 1986, *Knowledge of Language: Its Nature, Origin and Use*, Westport, CT: Praeger.

<sup>24</sup> My thanks to two anonymous referees for insightful comments. Also, my thanks to participants at a session of the 2014 Meeting of the Southern Society for Philosophy and Psychology where I presented some of the ideas in this paper.

- Chomsky, N. 1995, "Language and Nature", *Mind*, 104, 1-61.
- Chomsky, N. 2000, *New Horizons in the Study of Language and Mind*, New York: Cambridge University Press.
- Crane, T. 2013, *The Objects of Thought*, New York: Oxford University Press.
- Fodor, J. 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.
- Fodor, J. 1990, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.
- Fodor, J. 1998, *Concepts: Where Cognitive Science Went Wrong*, New York: Oxford University Press.
- Lau, J. and Deutsch, M. 2014, "Externalism About Mental Content", *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Zalta, E. (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/content-externalism/>
- Laurence, S. and Margolis, E. 2007, "The Ontology of Concepts: Abstract Objects or Mental Representations?", *Noûs*, 41, 561-93.
- Loar, B. 1988, "Social Content and Psychological Content", in Grimm, R. and Merrill, D. (eds.), *Contents of Thought*, University of Arizona Press, 99-110.
- Ludwig, K. 1996, "Singular Thought and the Cartesian Theory of Mind", *Noûs*, 30, 434-60.
- Machery, E. 2012, "Expertise and Intuitions about Reference", *Theoria*, 37, 37-54.
- McGinn, C. 1977, "Charity, Interpretation, and Belief", *Journal of Philosophy*, 74, 521-35.
- Parent, T. 2013, "Externalism and Self-Knowledge", *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Zalta, E. (ed.), URL = <http://plato.stanford.edu/archives/sum2013/entries/self-knowledge-externalism/>
- Rives, B. 2010, "Jerry Fodor", *Internet Encyclopedia of Philosophy*, URL = <http://www.iep.utm.edu/fodor/>
- Searle, J. 2004, *Mind: A Brief Introduction*, New York: Oxford University Press.