

## Can Fictional Superhuman Agents have Mental States?

Gabriel Levy  
Department of Philosophy and Religious Studies  
Norwegian University of Science and Technology  
NO-7491 Trondheim, Norway  
gabriel.levy@ntnu.no

### Abstract

*According to Deborah Tollefsen, from the analytic perspective called “interpretivism”, there is a reasonable way in which groups can be said to have mental states. She bases her argument on the every-day use of language, where people speak as if groups have states such as intentions, desires and wishes. Such propositional attitudes form the basis of any account of truth-conditional semantics, the rules by which people grasp the conditions under which an utterance is true. If groups (abstract units of people) have mental states, perhaps superhuman agents have them too. One argument that may contradict this premise is one that says that, whereas groups exist, superhuman agents do not. However, if groups exist on the basis of normative narratives about them and the institutionalized actions they carry out in the world, the same can be said for superhuman agents. They are like legal fictions: fictional but real. Superhuman agents are fictional and real in a similar sense as groups.<sup>1</sup>*

### Keywords

*religion, semantics, Davidson, fiction, truth, falsity*

---

<sup>1</sup> Thanks to Miriam Kyselo, Lionel Sacks, Ulrika Mårtensson, Rene van Woudenberg, the Department of Philosophy and Religious Studies at NTNU, John McGraw, Jeppe Sinding Jensen, Terry Godlove, Scott Davis, Lars Albinus, Nancy Frankenberry, and especially Mark Gardiner.

*Twitter exchange between comedian Norm Macdonald and Richard Dawkins:*

Norm Macdonald @normmacdonald 7/11/15  
 @RichardDawkins I was wondering if you thought it ever “reasonable” or “rational” for a human being to murder himself, or automatically mad.

Richard Dawkins @RichardDawkins 7/11/15  
 @normmacdonald This is a question for a psychologist, not an evolutionary biologist like me. If you’d lost all hope, family, career...?

Norm Macdonald @normmacdonald 7/11/15  
 @RichardDawkins I meant if you were perfectly healthy with a life that made you happy. Could a person simply opt for non-existence?

Norm Macdonald @normmacdonald 7/11/15  
 @RichardDawkins The reason I ask an evolutionary biologist is, would it be a necessarily mad act since it attacked self-survival?

Richard Dawkins @RichardDawkins 7/11/15  
 @normmacdonald From an evolutionary pov it can be rational. Gene survival matters, not individual survival. Look up ‘suicide’ in Selfish Gene.

Norm Macdonald @normmacdonald 7/11/15  
 @RichardDawkins I will. I read The Greatest Show on Earth and found you a gifted writer. Would you consider penning a non-fiction work?

Richard Dawkins @RichardDawkins 7/11/15  
 @normmacdonald Do you mean a fiction work? I think all my books so far have been non-fiction!

Norm Macdonald @normmacdonald 7/11/15  
 @RichardDawkins Sorry, yes, I meant fiction.

## **1. Introduction**

I am less concerned with making general arguments about Donald Davidson’s relevance to the study of religion as I am in applying his way of thinking concretely to religious language and see how far it gets us. In the course of his life Davidson eventually came to agree with Rorty’s

interpretation of Davidson's "view of truth" as neither a correspondence, coherence, nor pragmatic view of truth. In the course of writing his "Afterthoughts" (1987 (2001)) on "A Coherence Theory of Truth and Knowledge" (1983 (2001)), Davidson reiterated three points: first, with regard to correspondence, he wanted to expose "the false assumption that truth is transparently epistemic," that is, he argued that "truth is not an epistemic concept" (1987 (2001): 155). Second, with regard to coherence, he claimed that "coherence is nothing but consistency" (155). Third, he affirmed that he and Rorty were minimalists about truth (and falsity<sup>2</sup>) (155). Truth is about as basic a concept we can have, and can't be reduced to anything else; it is a primitive and "any further attempt to explain, define, analyze, or explicate the concept will be empty or wrong" (155). He sums things up succinctly: "Truth emerges not as wholly detached from belief (as a correspondence theory would make it) nor as dependent on human methods and powers of discovery (as epistemic theories of truth would make it). What saves truth from being 'radically non-epistemic' (in Putnam's words) is not that truth is epistemic but that belief, through its ties with meaning, is intrinsically veridical" (156). In other words, he was trying to characterize a holistic relation between meaning, belief, and truth—they come along together.

For Davidson, truth is an epistemic notion because, as he argues, it is tied to belief and belief is intrinsically veridical. He argues further that communication with language, even in its most basic form, cannot get off the ground without a background of massively shared true beliefs. This is a fundamental assumption of Davidson's holism. Such a version of holism opens up the problem of how we account for error and falsity. Just as truth can only be understood holistically, the same is true for error and falsity; thus Davidson contends that error cannot be met sentence by sentence, "even at the simplest level" (152); instead the "best we can do is cope

---

<sup>2</sup> Davidson paid far less attention to truth than falsity; but we can extend his minimalistic arguments about truth to the concept of falsity as well (see Levy 2012).

with error holistically, that is, we interpret so as to make an agent as intelligible as possible, given his actions, his utterances, and his place in the world” (152).

This perspective is the starting point of this essay, taking Davidson literally to be saying we *make* agents intelligible. *If* we want to understand them, know what they are saying (which we usually do), *then* we *make* them intelligible. When we interpret other agents charitably in this way, we are going to find that their beliefs are mostly true (and we are also going to find interesting areas of disagreement and error). It is in this sense that beliefs are by their nature veridical. Following this line, I argue that beliefs are also by their nature *fictive* in the etymological sense of the term, as something made. Names of fictional entities may lack referents, but they still have semantic content. The veridicality of beliefs about such agents is achieved by an act of imagination. In this respect interpretation and communication is more fictional than we usually think.

The challenge for this essay is to work out what this means within Davidson’s holistic framework. Since truth is not an epistemic concept, saying something is fiction or fictive is not the same as saying it is not true. Rather, fiction is agnostic about truth and falsity; it deals with more important matters. I argue that most communication is fictional in the sense that most books in the genre of non-fiction are fiction. Or rather, beliefs, when they bubble to the surface of reflection, when we talk about them, are fictional. They are fiction because no sense can be made of language corresponding to facts. But this does not mean that there is no such thing as objective reality.

I try to situate my use of the term fictional within the mediating position Davidson tries to work out between coherence and correspondence theories of truth. I will introduce a vocabulary for making sense of religious language that I think does justice to this point of view, but it

requires understanding the fictional not necessarily as false or disconnected from reality. Rather the term “fiction” is a way to avoid naïve forms of a correspondence account of truth, while retaining what Davidson sees as the intrinsic veridicality of belief. The big questions from the perspective of this approach will always be how we account for falsity, and whether attributing falsity to a specific set of beliefs (such as those of religion) really get us anywhere. These questions are not answered explicitly by Davidson, since he gives us no indication about *what* in fact makes things true or false.

Analytic philosophers of fiction usually opt for one of two views; either names of fictional characters, like gods (superhuman agents), unicorns and electrons, have no referents or else have referents in the worlds created by the fiction. I agree that such terms are fictional, but find it strange to say that these terms do not refer to *anything*. Perhaps the best we can do is call such terms “existentially challenged.” In the domain of existentially challenged agents from religious utterings, the gods, spirits, and beings of religion make most sense when thought of in relation to other agents produced in human narratives; as Salmon describes them, they are “fabricated but genuine” (2005: 101). I don’t think it necessary to say whether such agencies exist or not (for the purposes of understanding literature or religion). I think this is the closest a scholar of religion can get to the first-person perspective of a religious person—the content almost completely overlaps. That is, entertaining religious stories gets us a long way to understanding the first-person perspective, and in many ways, that is probably what religious people are doing as well. Fictional characters that we entertain are different in some respects than the superhuman agents in religions, and this gets us to the classical Mickey Mouse problem.<sup>3</sup> In

---

<sup>3</sup> See Pyysiainen (2003: 111) who, along with Scott Atran, first posed the issue of trying to understand the difference between fictional agents who are source of religious devotion and emotional commitment and those who are not. Pyysiainen’s hypothesis in that article,” is that

general the differences are more social than cognitive, though these domains overlap. I think there is a spectrum with regard to fictional characters. Some have huge social, political, and cultural apparatuses surrounding them, which support them and continue them, keeping them actively entertained in people's minds. The beliefs have real top down consequences for behavior and action. This is the case, for example, with Jesus, the Buddha, God, Shiva, and numerous other deities and spirits, including Santa Claus.<sup>4</sup> Others, such as Tom Sawyer, the BFG, or Calvin or Hobbes, do not have such an apparatus, aside from the material productions and perhaps some strange fan clubs.

## 2. Different types of information

I was talking to a friend who used to work as an academic at UCL in London and also at CERN; his training is mostly in electrical engineering and physics. I was telling him about an article in the *New Scientist* magazine (April 12-18, 2014) that I have been mentioning in recent talks. The article summarizes the work of MIT physicist Max Tegmark who argues that consciousness should be considered a basic state of the matter, along the model of solid, liquid, or gas. Part of what is going on there is a rethinking of what consciousness is—in this model it involves the way in which information is organized. I asked my friend whether this was a wild idea, and he

---

religious belief is possible partly because it exploits the evolved metarepresentational mechanisms that make the representation of fiction possible” (116). The article also broaches the very important subject of the relation between the concept of counterintuitivity and fiction (or even falsity). Accounting for the relation would connect my argument to the ongoing debate in the cognitive science of religion, but it is beyond the scope of this paper; see (Russell and Gobet 2013) and (Purzycki and Willard 2015).

<sup>4</sup> Justin Barrett (2008) disagrees, for he thinks Santa Claus does not qualify as a god, but I think he is wrong. The categories he uses to differentiate Santa Claus from other gods are arbitrary, *ad hoc*, and circular.

reminded me that it is basically a part of standard physics. The idea of entropy, he said, entails the concept of information within it, in the sense that entropy, loosely put, is a measure of order and disorder.

He was working for an IT company that builds some of the elements in the sensors in cars used to pay tolls. Entropy plays an even more important role in such contexts because it turns out that thermodynamic entropy—the kind studied by physicists regarding energy—can be understood in the same way, with the same equations, as the way information acts in telecommunication systems, based on theories of entropy and information first developed by Claude Shannon (1948). So, as in Tegmark's model, consciousness is a statistical measure of the amount of informational entropy. Nonliving things have very little of it, while living systems have much more of it. Plants have less of it than animals.

I like this line of thinking because I think it is a step in getting us out of what Hans Jonas called “the ontological dominance of death” (1966 (2001): 15) under which he thought modern scientific thinking was dominated. Jonas meant by this the idea that for modern science the universe is dead and science has to account for the emergence of life. This opposed a kind of Romantic archaic form of thought where the universe was alive and people had to account for death. I think Jonas wanted to go back to something closer to the latter form of thought, even though he did not think we could ever go back to it completely and he probably did not want to. Unlike Tegmark, Jonas would probably say that only living things have consciousness.

I think understanding the nature of information is crucial to understanding religion. I approach information from a perspective influenced by the philosophy of Donald Davidson, though it is an adaptation of his language. We can start with the idea that information is organized differently in different complexes. There is a subjective complex, an intersubjective

complex, and an objective complex. Subjectively organized complexes of information organize information around the *I* or first person. Another way to put this perhaps is that subjectively organized complexes are narratives about the *I*. Such narratives are arrived at in a different way than those at the intersubjective and objective levels. The former concerns the second person, and Davidson thinks this complex is what grounds the information at the other levels (subjective and objective) for human beings. The intersubjective level, in other words, represents what is special about the way information is organized in human beings. So Davidson favors the intersubjective, but we could just as easily say that all the levels are necessary to make sense of human information and communication. For Davidson, the level of objectivity is generated out of the level of the intersubjective. It is dependent on it.

This is not so crucial for my point except to say that I think religion can be viewed from all of these levels as long as we are careful in differentiating them when necessary; these different ways of organizing information are in some ways commensurable and in some ways not. Davidson, I think, would consider them anomalous—meaning they are commensurable in nonlinear ways.

### **3. Tollefsen's Interpretivism**

Deborah Tollefsen, in her book *Groups as Agents* (2015), glosses Davidson's approach in a very nice way. She makes it into a philosophical movement that she calls "interpretivism," which she says entails the view that, "if we can successfully make sense of another being—understand and interpret its behavior by using our folk psychology—it is an *intentional agent*" (2015: 113;



emphasis in original). She includes Haugeland (1990), Dennett (1987, 1991, 2011) and Mölder (2010) in this tradition (Tollefsen 2015: 113).

In a nutshell, this is an approach to intentionality “that starts not with metaphysical speculations about the nature of the mental, but with our practice of attributing intentional states,” where we ask “What assumptions do we need to make about an agent in order to interpret her behavior successfully?” “If interpretation is successful,” Tollefsen goes on to say, “then the assumptions we make about an agent in the process of interpreting her are justified” (2015: 113). The line between success and non-success is a pragmatic one concerning how far such fabrications get us in understanding the other person in the long run, a “pattern governed essentially by the norms of consistency and truth” (116).

In doing this, there is no need to look for further metaphysical facts. Mental states or intentions are not metaphysical states, they are ways of describing and organizing information about communication and interpretation; as Tollefsen says, they are states of “whole systems” (18, 115, 118, 126, 155). Taking interpretivism as a starting point has advantages because it gets us out of needless metaphysical speculation. However, if you disagree with interpretivism in this sense, namely if you are more concerned with the metaphysics of mental states (e.g. whether they exist or not or if we literally have them) than you are with their attribution and application, you may not find the argument that follows compelling. Nonetheless, casting the argument in terms of attribution rather than metaphysics is in line with the concerns of scholars of religion, who typically have social scientific interests that tend to outweigh their philosophical ones in this

regard—i.e. we are more concerned with the attribution of mental states than with their ontology.<sup>5</sup>

Tollefsen is aware that this approach, which sees successful understanding and interpretation as the master key, risks attributing intentionality to everything. She asks, “won't this mean that the class of intentional systems is quite large? ... We might explain the sedentary behavior of a flowerpot by saying that it wants to stay where it is and believes this is the best place for itself. Does this mean the flowerpot is an intentional agent?” (118). She thinks this would be an unhappy result but argues that there are various ways around it.

One such way is Davidson's, which says that in order to have intentional thought, one must have language, and further, to have human thought, one must be able to attribute not just intentions to agents, but also the minimal concepts of truth and falsity. In this sense, Davidson didn't think animals had full-fledged thoughts (1982 (2001)), although he did not derive any moral implications from this assertion. I think he meant animals didn't have thoughts in any sense we as human beings could recognize. Dennett's solution is wider, allowing for a range of intentional systems. What varies in attributing intentionality to such things as flowerpots, non-human animals, and human beings is the explanatory value of doing so. In the case of a flowerpot, attributing intentionality, so the story goes, does not get us a long way towards understanding it, because it “does not exhibit a complex pattern of behavior” (118). In other words, we don't need the concept of intentionality to make sense of a flowerpot. In sum,

---

<sup>5</sup> Not that I am unconcerned about their ontology. A major problem in cognitive science approaches to religion, for example, is that scholars tend to think of mental states as in the head. I don't think they are in the head, and I don't think we can really pull apart their attribution from their ontology. Mental states exist as long as they are needed and then they are gone (once they are gone we need to start talking about memory). If we follow Davidson, they are identical to physical states but not reducible to them. Davidson's holism incorporates mental states as part of theories (prior and passing); see (Davidson 1986).

rather than start with a list of criteria that a subject has to meet in order to be an intentional agent, interpretivism starts with our practice. It takes as its starting point the explanatory power of this practice. If we are able to understand and predict the behavior of a system using the intentional stance, then we have every reason to believe we are dealing with one (Tollefsen 2015: 119).

From here Tollefsen goes on to argue that groups can be considered agents with mental states.

#### 4. Toward an Interpretivist Approach to Fictional Superhuman Agents

One weekend while I was doing my usual reading of *The New York Times Book Review* (July 9, 2015 issue) I came across an article by George Soros entitled “A Partnership with China to Avoid World War.” Here is a typical quotation, in the context of Soros making a plea that China be included in the IMF (*italics added*):

The International Monetary Fund could play a positive part in this. It has abandoned its commitment to the Washington Consensus but the controlling shareholders of the Bretton Woods institutions—the US, the UK, France, and Germany among them—are *unwilling* to relinquish their voting control by increasing the representation of the developing world. This is very shortsighted on their part because it does not recognize changes in the relative weight of various economies and particularly the rise of China.

It is the most commonplace thing in the world that we speak of complex groups such as the US, China or Germany as agents with mental states, in this case *unwillingness*. Religious groups themselves might also be spoken about in this way. For example, this is from an article run in a newspaper in Philadelphia:

“The Archdiocese of Philadelphia, which is hosting the eighth triennial World Meeting in September, issued a statement Friday saying the investigation of Paglia did not appear to affect the meeting” (Philly.com “Archdiocese: Embezzlement inquiry has no effect on World Meeting” May 31, 2015).

In a nutshell, Tollefsen wants to ask whether, from the interpretivist perspective, we are justified in thinking of these and other groups as agents with mental states. These are not internal states, but attributional, dispositional states used to predict and interpret the behavior of complex units.

There are constraints on the attribution, namely, the group has to have some sort of unity—we don't take the intentional stance if the "agency is not unified enough," and furthermore the agency must be "extended over time in a way that makes their behavior complex enough to warrant the intentional stance" (Tollefsen 2015: 120). But ultimately, "if taking the intentional stance toward a group allows us usefully to understand the group's actions, then we have every reason to believe our assumptions of rationality are justified and that we are dealing with an intentional agent" (127). Such groups are organized and make decisions. Interpreting them in this way helps predict their actions; so for all intents and purposes, they *do* act.

I heard about this theory when Tollefsen gave a paper at a TESIS (Toward an Embodied Science of Intersubjectivity) conference in London in which she argued there is a reasonable way (at least from an analytic perspective) in which groups can have mental states. Afterwards, I asked her whether, if groups (abstract units of people) can have mental states, could perhaps superhuman agents have them too? She took a day to respond and said groups were different than gods because groups exist. I pushed her on this point—what does "exist" mean in this context?—but that answer was the best she could do at the time.

Tollefsen's is a strong rebuttal. There is a good case to be made that groups are real things that act in the world. But I think in making this move, she commits an error she accused others of; namely moving from a metaphysical theory to an interpretive theory rather than the other way around. People attribute actions to groups in narratives, such as the one above by

George Soros. Such narratives help people like Soros make predictions about complex things like the global financial and political systems. The same perhaps is true of fictional superhuman agents, or at least that is the idea I would like to defend.

As someone trained in Biblical scholarship, I can't help but point out that much of the early mythology was exactly this kind of narrative—trying to make sense of groups of people either named with the area, ruler, or god. For example, from the book of *Numbers* (20:21):

So Edom would not let Israel cross their territory, and Israel turned away from them.  
(JPS)

וַיִּמָּאֵן אֱדוֹם וְלֹא יָרְאָה אֶת-יִשְׂרָאֵל לְעָבַר בְּגִבְלוֹ וַיִּטּוּ יִשְׂרָאֵל מֵעָלָיו:

Language like this is ubiquitous in the Bible; rather than get bogged down in what exists, let's stick with interpretivism and consider whether attributing mental states to fictional superhuman agents has explanatory value.

## 5. Varieties of Fiction

I delivered a version of this article for a group of philosophers in the “Science Beyond Scientism” project at the VU University in Amsterdam. The leader of the group, Rene van Woudenberg, had a similar easy answer to the question posed in my title: “Can fictional agents have mental states?” *Clearly no*, since fictional things don't exist, and only things that exist can have mental states. He brought up Frege's notion of existence and non-existence, namely that there is some greater than zero number of things that exist, but only zero number of things that do not exist. Things that there are zero of can't have mental states. But then again, Rene is a

theist, so he thinks there is at least one god that exists in his Fregean sense of the word, and thus is not fictional.

These are the type of troubles one gets into in calling something fictional. As mentioned at the outset, I mean *fictional* or *fictive* in a slightly unconventional way, more about which will be said subsequently. In answer to the Fregean point, I think it obvious that fictional agents are of a certain number that is not zero. In other words, there are non-zero numbers of Huckleberry Finns and Billy Pilgrims (this is the idea many authors explore, including Kurt Vonnegut in his last novel *Timequake* from 1997).<sup>6</sup>

I am arguing that religions can be thought of as communal systems of propositional attitudes and practices related to fictional superhuman agents. Since, within my framework all superhuman agents are fictional, given recent trends in theorizing about religion (see note 3) we can push this to say religions are communal systems of propositional attitudes and practices related to *fictional* agents. If the analogy with Tollefsen's groups is correct, then sometimes we may be justified when we attribute mental states to such agents.

Fictional characters are alive in the minds of both writers and readers. Serious novelists often report that, at a certain point in the process, their characters take on a life of their own.

---

<sup>6</sup> If I am grasping Van Woudenberg's oral criticism of my paper correctly, he seemed to be making a simple point: that fictional agents don't exist, so of course they cannot have mental states. However, he may not have got Frege right in his criticism. At least according to one expert, "Frege himself had very little to say about nonexistent objects" (Parsons 1982). Van Woudenberg was enlisting Frege to argue that the set of fictional entities is, in fact, the empty set. But, sets are always relative to a domain or universe of discourse, at least in standard set theory. There is a set-theoretic sense in which Huckleberry Finn does not exist (i.e. he is not in the set of objects, say, that one can empirically observe) and another in which he does exist (i.e. he is in the set of objects, say, that interact with Tom Sawyer). Thanks to Mark Gardiner for help with these points. There is an ongoing line of inquiry among analytic philosophers concerning fictional and non-existent entities. For a recent book-length analysis, see Sainsbury (2010), who presents three different arguments about the reality of fictional objects: (i) they are real but nonexistent (chapter three); (ii) real but nonactual (chapter four); or (iii) real but nonconcrete (chapter five).

They feel like they are no longer in control of what their characters say and do, with the novelist simply reporting what happens. For readers, fictional characters are treated just like non-fictional characters that we read about. Fictional agents are real organizations of information. We reason about them. Dennett goes so far as to argue that, in some sense, the self is such an agent: “A self is... an abstract object, a theorist’s fiction,” like “center of gravity” in physics (1992: 103). Philosophers of language debate whether such fictional names refer to anything, e.g. whether they have a regular or “gappy” semantics, but few doubt that they still function semantically and inferentially.<sup>7</sup> A further question is whether second-order discourses about fictional stories, rather than the stories themselves, are also fictional. In other words, can we distinguish the semantics within the narrative from the semantics of discussion about the narrative?

Religious discourse seems intriguingly similar; Davis’ discussion of midrash (this volume, pp. \*71-81\*) provides a nice example. Fan fiction may also provide a convenient parallel, where fictional agents continue to live even beyond the works of the original authors. Religions are information complexes where the attribution of agency makes sense. Durkheim was probably the first to understand that language about religion was derived from language about groups. He looked at this from an evolutionary perspective, but we could also look at it from a developmental perspective. That is, children learn to tell stories about, or at least attribute intentions, to agents they can’t see—e.g. they quickly pick up on, without explicit instruction, how adults use “China” to name an intentional agent.

My main argument thus rests on accepting an interpretivist account along the lines of Tollefsen about groups, while less of it rides on the question of whether or not we consider

---

<sup>7</sup> For background on the issue of fictional names, see Adams, Fuller, and Stecker (1997), Eco (2009), Goodman (2013), Krasner (2002), Salis (2013), Thomasson (2003), Vacarescu (2015), and Salmon (2005).

superhuman agents fictional. However, I think this question is important. For example, it provides nuance to Hans Penner's claim that religious language is patently false (1995; see also Levy 2012); regarding religious language as a kind of believed fiction is the best way to understand Penner's claim. The question is also important in that it will allow us, I will argue below, to think of scientific, not just religious, language as a form of fiction.

Penner's argument about the patent falsity of religious language has been underappreciated. Perhaps this is because he is simply wrong, but I think it is more likely because we haven't done an adequate job of thinking through his argument. He had good reasons to make the claim, based mainly on Davidson's "What metaphors mean" (1978 (1984)). Davidson argued that metaphors are patently false; there is a difference between saying that Tolstoy is *like* a moralizing infant (which is a simile and, like all similes, is patently true as everything is like anything else in some respect or other) and saying that Tolstoy, as the author, *is* a moralizing infant (which is a metaphor, and patently false) (1978 (1984): 247-8).<sup>8</sup> It might be the case that the analogy between religion and metaphor breaks down and so Davidson's argument about metaphor does not apply to religion, or that Davidson's original argument about metaphor is faulty in some respect. However, assuming both are correct, I want to explore the implications of Penner's argument through the lens of interpretivism.

---

<sup>8</sup> Davidson does not disparage the usefulness of metaphor—rather he thinks that it has the use that it does precisely because its patent falsehood jars us into thinking about what real point is that the speaker is making (see Levy 2012).



## 6. False vs. True Fiction

Penner says religious language is patently false. We know from a holistic Davidsonian perspective that we can't pick out false sentences one by one and we can't even decide the rules for what generally makes one sentence true and another one false; as he argues, "Nothing, however, no *thing*, makes sentences and theories true: not experience, not surface irritations, not the world, can make a sentence true" (Davidson 1974 (1984): 194). If we are right about this, then how can we account for relatively massive disagreement, such as the case with religious beliefs, from a holistic perspective? We can't take each "religious" belief one by one, but perhaps we can talk of the difference between religious institutions or systems and scientific ones.

The first point to get out of the way is that science and religion *do not* constitute separate worldviews. The very idea of worldview is based on a dualistic conception of language. In other words, religion and science are not differentiated because they are two systems with different relativistic worldviews; they are so not differentiated because no sense can be made of the idea of worldviews that would be radically untranslatable from one to the other. Hence, "science" *must* come into conflict with "religion". But what does this mean?

Part of Davidson's form of monism was a strong critique he mounted on what he called, following Quine, "a dualism of total scheme (or language) and uninterpreted content," (Davidson 1974 (1984): 187). Such a dualism radically distinguishes between conceptual schemes and the content people claim those schemes organize. Davidson thought no sense could be made of such an idea. More recently scholars in the study of religion have taken hold of Davidson's argument to criticize approaches to religion that smack of the scheme-content dualism. For example, Frankeberry and Penner, in their critique of Clifford Geertz, argue that those who hold on to the

dualism believe they can stand outside the fray of symbol making and posit simple correspondence or non-correspondence with the world. As they put it:

Conceptual schemers, as we might call them, hold that there are widely different “modes of seeing” the world, different perspectives, such as science, or religion, or common sense, which use different “symbolic vehicles” to signify what the world is all about. Thus different languages or symbolic vehicles cope with, organize, or “see” reality in importantly different ways. Symbol systems serve as colored glasses, or filters, or screens through which one's knowledge of the world, or the resolution of human problems with the world, must pass. Such perspectives are presented as incommensurable where incommensurability means untranslatability (1999: 632).

Following Davidson, they argue that such dualistic views are not intelligible. If the anti-scheme/content dualism scholars (the conceptual non-schemers) are right, no sense can be made of radically distinct worldviews. As far as it applies to the topic at hand, this means that it is a mistake to see science and religion as worldviews. As Frankenberry and Penner note in the previous sentence to the paragraph above, “Together with the association of the religious perspective with a sphere one can only enter by a leap of faith, this claim about the impossibility of even asking the question concerning the truth of religious beliefs links Geertz with a large group of scholars of religion who view religion and science as two different conceptual systems” (632).

So from the point of view of this monistic criticism, the concept of worldview has been discredited; the only sense that can be made of the concept is that every person has a view of the world, but if there are as many worldviews as there are people, the concept does not make sense—it doesn't do the work it was put there to do.

If there are no such things as worldviews, how can we imagine science and religion coming into conflict, or not coming into conflict? The answer is that it depends what we mean by the terms science and religion. From this monistic perspective, while science and religion do come into conflict—both are constructions, but not all constructions are made equally—it is not

in the places we usually think they do. There will be conflicts, but also surprising and unexpected areas of confluence. These areas of confluence come out especially when considering, for example, the relations between “mysticism” in the Late Medieval *La Convivencia* sense between Abrahamic religions, Eastern, and Greek thought.<sup>9</sup> While the concept of *La Convivencia* may be exaggerated because it did not apply to the vast majority of people living in Iberia in the middle ages, it probably did apply to many intellectual, political and religious elites. Magic, science, and religion were not yet differentiated, as Tambiah (1990) argued. This differentiation did not happen until after the Reformation. Before this time, “magic and science. . . advanced side by side and. . . mystical-magical theories and preoccupations advanced the formulation of those theoretical systems that would later be seen as triumphs of the new science: examples are heliocentrism, the infinity of the worlds. . . the circulation of the blood, and certain applications to mathematics” (Tambiah 1990: 21).

Tambiah tried to see through the dichotomizing tendencies of what he called “Whig” historians of science, who draw linear historical lines, by telling the stories of a few of the many earlier “scientists” whose interests blurred the boundary between modern conceptions of magic, science, and religion, such as Giordano Bruno, Marsilio Ficino, John Dee, in addition to Copernicus and Newton, to name a few. During the Reformation and until recently a narrative emerged that there was a radical break, a Kuhnian revolution. While there has been some differentiation, particularly in the way science is in principle “a game without end. . . its knowledge is regarded as provisional, and it accepts rules by which one theory is to be replaced

---

<sup>9</sup> Which I argue, following Tambiah (1990), was an important link in the chain provoking the scientific revolution. See, for example, Harris (2015), where he and Max Tegmark discuss cosmology.

by another if they are in competition,” (142) Tambiah thinks it a bit overblown.<sup>10</sup> He notes, thinking of astro- and theoretical physics in particular, that present sensibilities are more tolerant “of the possibility that advance in the application of science can live happily with paper and laboratory manipulations of imagined, even metaphorical and fanciful cosmologies and mathematical explorations” (29; see also Harris 2015).

Human and natural sciences rose out of the first European universities, which themselves originated often as monasteries or other types of religious institutions. Following Tambiah, we can see that the sometimes-justified distinction between human and natural sciences is probably a Protestant binary. The original religious institutions were based on metaphysical biases in Christianity, especially in its Protestant forms, and also perhaps on what I have heard described as “protestant forms of Judaism,” such as offered by Spinoza (see Hunter 2005). I do not intend these points directly as criticism, because all education systems are informed by the theological and metaphysical binaries that surround them. As for the European university, it is not really until the twentieth century that other theologies and metaphysics become part of the university, most visibly in departments of religious studies.

I find it ironic that as soon as a non-Christian viewpoint is made possible in the university, it is at this moment that people want to get rid of the humanities, that we have a major so-called “crisis” in the humanities. There is no room for science and non-Christian religion to remain somehow together.

---

<sup>10</sup> Consider where he characterizes the difference between science and magic in the way that the latter “lacked a secular and critical attitude that is the guarantee of open-ended scientific knowledge” (31). Tambiah’s invocation of the idea of “secular” in this passage begs the point. For Tambiah, the scientist is “in a special way open to the provisional nature of his knowledge” (142).

I have a model for how they could. In this model both science and religion are *fictions*. Religion is a special type of fiction, and so is science. One is true fiction, one is false fiction. They are fictions in the same way that the story Soros tells about the USA and China are fictions, or novels are fictions, corporations considered as legal agents are fictions, or the way other legal fictions are fictions. Both science and religion are narratives that organize information. But they organize it in different ways.

I derive this idea from Penner's use of Davidson on metaphor to argue that religious language is patently false. Davidson understands truth and falsity from a semantic perspective. They are irreducible primitive operators, akin to a kind of grammar of semantics, which provide content to communication. People often get bogged down by the term "truth". As noted, Davidson didn't think truth involved either correspondence (having to do with a relation between sentences and the world) or coherence (having to do with the way sentences relate to one another) but some strange mixture of the two (see Davidson 1983 (2001)).

Since most of our basic beliefs are true, according to Davidson, and since, he further argues, successful communication is dependent on a kind of holistic score-keeping (like a game with rules) where we must agree by and large where our truths and falsities overlap (for otherwise the communication would not be successful; the game wouldn't work), we don't have to worry so much about the individual cases. Without the flutter between truth and falsity, communication would not be possible. There is ongoing debate in the study of religion about how useful a truth-conditional approach to religion actually is (that is what this volume is about); but we can say, in his use of Davidson, Penner was one of the first to give us an interpretivist theory (in the Tollefsen sense) of religion.

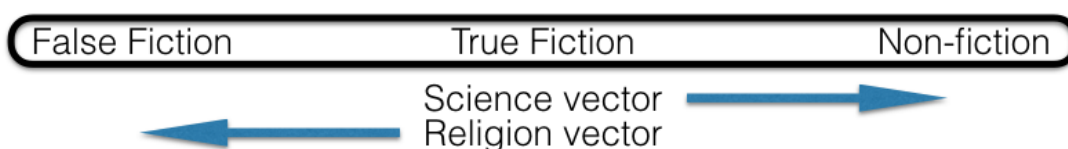
However, before we go too far after religion, we must say also that most *scientific* language is *also* false. As I will try to explain below, most science comes to us in the form of narratives that are in fact false. This is especially true when “scientists” discuss “religion”.

This is not a relativistic point. Science is a very good way to discover truths and falsities. They often come out in the scientific process. But the true part of science is a pretty tightly bound system of information that doesn’t really get out of its experimental context, the particular language game where it makes the most sense (see Levy 2014: chapter five). This is another way of asserting Davidson’s notion that objectivity is dependent on intersubjectivity. Objective systems are not communicative.

Most science one runs into, in the form of magazines like *New Scientist*, “science” shows on TV, references to it by people like Richard Dawkins, is fictional. In the twitter conversation between the Canadian comedian Norm McDonald and Richard Dawkins pasted at the beginning of this chapter, Norm should have stuck to his guns. Richard Dawkins writes mostly fiction. The only non-fictions he writes are scientific articles in his area of specialty about insect biology. There is some science behind his popular books, but the actual output is driven by interests more akin to myth. As Leonard Mlodinow points out in his recent book *The Upright Thinkers - The Human Journey from Living in Trees to Understanding the Cosmos* (2015), talk in science is full of mythical narratives—usually on the model of revelation—such as Newton discovering gravity when an apple fell from a tree or Darwin discovering evolution on the Galapagos islands. These myths fit a kind of lone genius pattern that ignores the collective basis of scientific development.

But it is wrong to think of the relation between fiction and non-fiction as a binary; it is better to think of it as a continuum. When I say science is true fiction, I am saying most scientific discourse is fictional, but much of it is a hybrid between fiction and non-fiction. This

conceptualization can be represented on a scale from true fiction to false fiction to non-fiction. As represented in Figure 1, science vectors toward non-fiction, while religion vectors toward false fiction.



*Figure 1: Vectors of Religion and Science*

This leaves us the question of where the science is in science? I think it is pretty clear that the science in science can also be found outside science, and it is found in arts of self-correction. Modern science has rigorous methods to do this, and technological means to repeat and store the outputs of these corrections. So, it is in this sense that science is true fiction.

To reiterate: religion and science are both genres of fiction. Two points follow out of this that I want to focus on.

The first is that religion is false fiction—it doesn't do much or any self-correcting; at least it doesn't have a rigorous method for that. Of course, this point depends what we mean by religion. Religion may do other things of value, and some religions might even have methods for self-correction; but even if it does, the time-scale for doing so is vastly slower than science.

The second point is that as far as science or true fiction is concerned, we humans spend only a small part of life self-correcting. Even scientists probably don't do much of it, except within very small confines. In other words, the true part of true fiction is only a small part. The rest of science is filled with cultural (think of the military industrial complex, advertising science

that urges consumption, or the pursuit of money), psychological (cheating to get ahead and make a name for oneself), and religious (think mind-body dualism) constraints.

Another way to say this may come from a first-person perspective; we can say that true fiction is based on operationalizing ongoing theories, subjecting them to experiments or other efforts that test the theories based on the operationalization, deriving results for, against, or somewhere in between, and then saying things *I* believe to be true and derivable from those results. In terms of false fiction, the *I* is also saying things believed true, (depending on one's definition), such as things believed true about superhuman agents. The things I believe true are not derived from operationalizing ongoing theories, subjecting them to experiments or other efforts that test the theories based on the operationalization.

I think the traditional methods of the humanities are more like religion: reading, being critical, and saying things I believe to be true. I would like to see a new humanities, one that is informed by the Mind Sciences, involving a hybrid of these methods: conducting "experiments", reading, deriving results, saying things I believe to be true and derived from those results.

To get back to the main point, we can say that the outputs of science and religion are fictional. This does not mean that the things they refer to don't exist. As Quine famously put it:

As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries—not by definition in terms of experience, but simply as irreducible posits *comparable, epistemologically, to the gods of Homer*. For my part I do, qua lay physicist, believe in physical objects and not in Homer's gods; and I consider it a scientific error to believe otherwise. But in point of epistemological footing the physical objects and the gods differ only in degree and not in kind. . . .

Positing does not stop with macroscopic physical objects . . . .

Physical objects, small and large, are not the only posits. Forces are another example; and indeed we are told nowadays that the boundary between energy and matter is



obsolete. . . . *Epistemologically these are myths on the same footing with physical objects and gods*, neither better nor worse except for differences in the degree to which they expedite our dealings with sense experiences (1951 (1953): 44-5; emphasis added).

There needn't, therefore, be any tension in both saying that superhuman agents are fictional and that they exist. We don't need to go beyond the cognitive and linguistic level to make sense of language about them.

## 7. Agents in Flat Ontology

So what about Tollefsen's retort that fictional superhuman agents don't exist and so aren't analyzable through the lens of interpretivism? I think this is the outcome of someone who doesn't know much about religion. It is no fault of her own; she has other concerns. She is pushed in that direction not by the contours of her interpretivism but by other uncritical ideas about what religion is.

Aligned with my rebuttal already presented above, a different philosophical tradition, actor-network theory and "object oriented ontology" (Bialecki 2014), converges on a similar point. The convergence of distinct scholarly traditions is indeed a happy coincidence of ideas. The term "object" is unfortunately misleading however, for, according to Bialecki, the theory states that all objects are potential "actants" (so are not really objects in the traditional passive sense of the term, but this is precisely the point). He argues that the world is not composed of "subjects on one hand and noumenal objects on the other," but rather, "of nothing but objects, animate and inanimate, human and nonhuman, all of which have to be taken as agents (occasionally glossed as "actants")" (35). Such an account entails what Bialecki calls a "flat ontology," (38) in the sense that if something looks like an object/actor and quacks like an

object/actor, it is one. The ontology is flat in the sense that all objects are created equal.

Similarly, I think a flat ontology is one of the more profound implications of interpretivism.

Bialecki combines Webb Keane's and Bruno Latour's ideas about religion as a starting point. Latour and Keane point to the centrality of "purification," which is "the continuing attempt to disarticulate nature from culture, human agency from dross materialism" (37). I take purification to mean the drive to purify nature of agency; on this reading, agency and associated concepts like mental states belong to human individuals, and any attempt to attribute them to other entities introduces semantic impurity that is perhaps best dealt with as metaphor.

Such an account, even from an atheist perspective, allows us to speak of "God" as a hybrid—both object and actor. The concept of "God" is built up and composed of a series of different object/actors. In the ethnographic example of *The Vineyard*, a Pentecostal Christian movement in the USA, people speak to God, even setting a coffee cup and empty chair up for Him to sit in.<sup>11</sup> The iterated performances of this interaction enact and create God the actor. The hybridity—where God is "built of unconnected heterogenous objects" such as chairs, coffee cups and performances—is, in effect, what allows God to become a real social actor. The heterogenous composition "allows God to function as a constant source of surprise" (Bialecki 2014: 42). This exceeding of expectations and surprise creates a fluttering "sensation that they are interacting with an autonomous being with an independent will" (42).

Perhaps we can augment Tollefsen with this understanding of fictional agents. Agency and the attribution of mental states is not just about predicting behavior, but also *not* being able to predict such behavior in terms of the sense of surprise noted by Bialecki. In this sense,

---

<sup>11</sup> The performance sounds conceptually similar to Clint Eastwood's fictional interview of Barack Obama at the 2012 Republican National Convention. In this example, Obama is analogous to God in the *Vineyard*. See <https://www.youtube.com/watch?v=933hKyKNPFQ>.

fictional superhuman agents are much more like human agents (who are also fictional in some sense, see (Dennett 1992)) than we think, for “God acts with the same kind of stochastic wildness that is usually allocated to the other human agents who stand in the center of so much of contemporary anthropological writing” (Bialecki 2014: 42-3). Bialecki thus concludes that “in societies where He is produced, we have to ethnographically include God—and specifically, include him as a (potential) social actor” (38). Though I think it unclear exactly what Bialecki means by this inclusion, the idea of ethnographically including God or gods as potential social actors converges on my point.

## **8. Conclusion**

I endorse this idea of flat ontology—it allows for a form of atheism that at the same time accepts that characters like “God” can be agents. To say one can be an atheist but still think that superhuman agents like gods, demons, buddhas, and spirits exist does seem to change around the normal use of words, but I have argued that the position is defensible along the lines of Davidson’s point of view that does not commit itself to correspondence or coherence theories of truth. If the content of language is tied to a view of truth as a twisted mix of correspondence and coherence, if nothing in the world makes a sentence true or false, if language is deeply embedded in the world and is radically contextual, and if our beliefs are massively veridical, then talk of existence is usually a misguided form of talking. Fictional characters that come alive in everyday language have the potential to be agents, some more so than others. I consider this still a form of atheism because I characterize superhuman agents as fictional—but it turns out not much rides

on this particular designation, other than signaling a resistance to relativism (see Figure 1) and linguistically naïve forms of metaphysical speculation.

The issue of avoiding correspondence theories of truth is what makes essays like “What metaphors mean” (Davidson 1978 (1984)) and Hans Penner’s use of it (1995; Levy 2012) especially strange. In these essays, Davidson and Penner come down to decide that some ways of speaking are specifically false. They do not say why—and Davidson advises us that such an account would not be possible from a semantic point of view. But we are nevertheless told that some types of speaking *are* false. This is why Gardiner and Engler’s (2008) critique that Penner’s argument smacks of a correspondence theory of truth (and falsity) is relevant (see Levy 2012: 150). Penner seems to claim that there is something essential about metaphors or religious language—some lack of correspondence between what the words literally mean and what they are applied to—that makes them false.

In the same way that Davidson attributes falsity to metaphor, I am attributing fictionality to science and religion—making no metaphysics or correspondence claims in the process. In other words, they are not fictional because of some relation they have or do not have to the world. I am trying to save Penner (and Davidson) to some extent from the accusation of smuggling in a correspondence account of truth with this nomenclature. The designation “fiction” allows us to remain agnostic on metaphysical claims, because in many ways “fiction” is just as real, or perhaps even more real, than “fact.” As is said, “the truth is stranger than fiction.”

What are the main implications of this argument for method and theory in the study of religion? Theoretically, the main takeaway presents the semantics of religion along the lines of gappy fictionalism and interpretivism: religions involve fictional agents whom we often relate to by justifiably attributing to them mental states. These agents are real in a sense that is different

than the reality usually attributed to them by theists. Nonetheless the position allows for greater agreement with the theist, and thus does not require bankrupt scheme-content dichotomies like “emic” and “etic” perspective-taking; neither does it require accusations of irrationality or counter-intuitiveness.

From a methodological perspective, a few consequences follow. Scholars of religion should utilize a mixed methods approach to investigate the parallels between literary fiction and religion. While presumably there are some important differences between purely fictional systems and religious ones, these can be investigated from both cognitive and social perspectives, which are not mutually exclusive. Scholars of religion, especially cognitive scientists of religion, must take seriously the dimensions in which religious imagination makes sense, even in its falsity—especially in its falsity.

## Bibliography

- Adams, Fred, Gary Fuller and Robert Stecker. (1997). The semantics of fictional names. *Pacific Philosophical Quarterly* 78 (2): 128.
- Barrett, Justin L. (2008). Why Santa Claus is not a god. *Journal of Cognition & Culture* 8 (1/2): 149-161.
- Bialecki, Jon. (2014). Does God exist in methodological atheism? On Tanya Lurhmann's when God talks back and Bruno Latour. *Anthropology of Consciousness* 25 (1): 32-52.
- Davidson, Donald. (1974 (1984)). On the very idea of a conceptual scheme. In *Inquiries into Truth and Interpretation*, 183-198. Oxford: Clarendon Press.
- .(1978 (1984)). What metaphors mean. In *Inquiries into Truth and Interpretation*, 245 - 264. Oxford: Clarendon Press.
- (1982 (2001)). Rational animals. In *Subjective, Intersubjective, Objective*, 95-105. Oxford: Oxford University Press.
- (1983 (2001)). A coherence theory of truth and knowledge. In *Subjective, Intersubjective, Objective*, 137-153. Oxford: Clarendon Press.
- (1986). A nice derangement of epitaphs. In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, 433-46. Oxford: Blackwell.
- (1987 (2001)). Afterthoughts. In *Subjective, Intersubjective, Objective*, 154-157. Oxford: Clarendon Press.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- (1991). Real patterns. *The Journal of Philosophy* 88 (1): 27-51.
- (1992). The self as a center of narrative gravity. In F.S. Kessel, P.M. Cole and D.L Johnson (eds.), *Self and Consciousness: Multiple Perspectives*, 103-115. Hillsdale, N.J.: L. Erlbaum.
- (2011). Intentional systems theory. In B.P. McLaughlin, A. Beckermann, and S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind*, 339-350. Oxford: Oxford University Press.
- Eco, Umberto. (2009). On the ontology of fictional characters: a semiotic approach. *Kirjanduslike Kangelaste Ontoloogias: Semiootiline Lähenemine*. 37 (1/2): 82-98.

- Frankenberry, N K, and H.H. Penner. 1999. Clifford Geertz's long-lasting moods, motivations, and metaphysical conceptions. *The Journal of Religion* 79 (4): 617–40.
- Gardiner, Mark Q., and Steven Engler. (2008). Review essay: Nancy K. Frankenberry (ed.), *Radical Interpretation in Religion. Method & Theory in the Study of Religion* 20 (2): 185-190.
- Goodman, Jeffrey. (2013). Creatures of fiction, objects of myth. *Analysis* 74 (1): 35-40.
- Harris, Sam. (2015). The multiverse & you (& you & you & you...): a conversation with Max Tegmark. In *Waking Up*. September 23, 2015. <https://www.samharris.org/podcast/item/the-multiverse-you-you-you-you>.
- Haugeland, John. (1990). The intentionality all-stars. *Philosophical Perspectives* 4: 383-427.
- Hunter, Graeme. 2005. *Radical Protestantism in Spinoza's Thought*. Aldershot: Ashgate.
- Jonas, Hans. (1966 (2001)). *The Phenomenon of Life; Toward a Philosophical Biology*. Evanston, Illinois: Northwestern University Press.
- Krasner, Daniel A. (2002). Semantics and fiction. *Erkenntnis (1975-)* 57 (2): 259-275.
- Levy, Gabriel. (2012). False but significant: the development of falsity in religious cognition in light of the holism of the mental. *Method & Theory in the Study of Religion* 24 (2): 143-165.
- (2014). *Judaic Technologies of the Word*. New York: Routledge.
- Mlodinow, Leonard. (2015). *The Upright Thinkers: The Human Journey from Living in Trees to Understanding the Cosmos*. New York: Pantheon Books.
- Mölder, Bruno. (2010). *Mind Ascribed: An Elaboration and Defence of Interpretivism, Advances in Consciousness Research (Aicr)*. Amsterdam: John Benjamins Pub. Co.
- Parsons, Terence. (1982). Fregean theories of fictional objects. *Topoi* 1 (1-2): 81–87.
- Penner, Hans H. (1995). Why does semantics matter to the study of religion? *Method & Theory in the Study of Religion* 7 (3): 221- 249.
- Purzycki, Benjamin Grant, and Aiyana K. Willard. (2015). MCI theory: a critical discussion. *Religion, Brain & Behavior* 6 (3): 207–48.
- Pyysiainen, Ilkka. (2003). True fiction: philosophy and psychology of religious belief. *Philosophical Psychology* 16 (3): 109-125.
- Quine, W. V. (1951 (1953)). Two dogmas of empiricism. In *From a Logical Point of View*, 20-46. Cambridge: Harvard University Press.

- Russell, Yvan I, and Fernand Gobet. (2013). What Is counterintuitive? Religious cognition and natural expectation. *Review of Philosophy and Psychology* 4 (4): 715–49.
- Sainsbury, R. M. (2010). *Fiction and Fictionalism*. New York: Routledge.
- Salis, Fiora. (2013). Fictional names and the problem of intersubjective identification. *Dialectica* 67 (3): 283-301.
- Salmon, Nathan U. (2005). *Metaphysics, Mathematics, and Meaning*. Oxford: Oxford University Press.
- Shannon, Claude E. (1948). A mathematical theory of communication. *Bell System Technical Journal*. 27 (3): 379–423.
- Tambiah, Stanley. 1990. *Magic, Science and Religion and the Scope of Rationality*. Cambridge: Cambridge University Press.
- Thomasson, Amie L. (2003). Speaking of fictional characters. *Dialectica* 57 (2): 205-223.
- Tollefsen, Deborah. (2015). *Groups as Agents*. Cambridge, UK: Polity Press. Accessed as E-book with Google Play (John Wiley & Sons Copyright).
- Vacarescu, Theodora-Eliza. (2015). What's in a name? Modest considerations on the situatedness of language and meaning. *Journal for the Study of Religions and Ideologies* 3 (9): 124-135.
- Vonnegut, Kurt. (1997). *Timequake*. New York: G. P. Putnam's Sons.