Non-Self-Exemplification Exists!

Abstract

‘Non-self-exemplification’ expresses a paradoxical property that exemplifies itself if and only if it doesn’t. The standard view is that the predicate fails to express a property and that we must therefore pursue a *restricted* theory of properties. I show, however, that the argument against non-self-exemplification implicitly depends upon a non-trivial, semantic assumption. I then make use of a theory of *having* and *lacking* properties to show how non-self-exemplification may *lack* itself without thereby exemplifying its complement. This theory empowers an unrestricted theory of properties and, as a bonus, provides a foundation for a powerful, “positive” theory of sets.

Non-Self-Exemplification Exists

"The paradox of non-self-exemplification is curled like a worm in the very heart of human reason; perhaps it was put there to keep us humble." – Alvin Plantinga[1]

1. The Problem of Non-Self-Exemplification

Some properties exemplify themselves: for example, *being abstract* exemplifies *being abstract*. These properties are characterized, then, by *self-exemplification*. Others aren't: *being brown*, for example, lacks self-exemplification. We may say, then, that while some properties are characterized by self-exemplification, others are characterized by a *lacking of self-exemplification*.

But there is a famous problem with supposing that there is any such property as *lacking self-exemplification*. The problem is that the property itself both *does* and *does not* exemplify itself: (i) it does exemplify itself because if it didn't, then it would exemplify its complement, which is *self*-exemplification; and (ii) it doesn't exemplify itself because if it did, then it would *lack* self-exemplification. So, to avoid the contradiction, we must reject the existence of such a property.

The problem isn't just that there couldn't be a certain *property*. There is a problem here for nominalists, too. Consider the *predicate* 'lacking self-exemplification'. This predicate is perfectly intelligible: you know exactly what I mean when I say that the Eiffel Tower lacks self-exemplification. Don't you? If so, then you should also understand the question, "Does the predicate, 'non-self-exemplification' apply to itself?" The problem here is that the only possible answer appears to be "yes *iff* no", which is absurd. The problem of non-self-exemplification is a problem for the metaphysician and the philosopher of language. [2]

I have not found any remotely satisfying account of why 'non-self-exemplification' should fail to express a property or to be a meaningful predicate. There was a time when I suspected the problem may have to do with the term "not": maybe there are no "negative" properties. But I couldn't convince myself that there are no meaningful negative *predicates*. I say, "I am not hungry," and people know what I mean.[3] Then there was a time when I thought maybe the problem is with the term "self" in "self-exemplification". But upon reflection, it seems there is no good way to motivate the suggestion that "self" is problematic *in general*. There is no conceptual problem, for example, with supposing that every boy on the bus is thinking about

[1] Plantinga said this in a lecture some time ago at the University of Notre Dame. I confirmed the quote via e-mail, May 5th, 2013.

[2] Hence, there is the famous Grelling-Nelson paradox, which is basically the "philosophy of language" version of the paradox of non-self-exemplification.

[3] I don't believe it will help to treat 'not' as a sentential operator. If we translate "I am not hungry" as "not, I am hungry," then what this sentence says is that the *proposition that I am hungry* is *not true*..

himself.[4] And if there is no problem in general, then I am left wondering why “self” should pose a problem in particular cases. There is also the “levels” approach, which attempts to escape paradox by supposing that a property is only exemplified by entities a level down a hierarchy.[5] But levels in this context are difficult to define. Is *being abstract* below *being a property*, or vice versa? In either case, *being a property* fails to exemplify *being a property*, and *being abstract* fails to exemplify *being abstract*. But both are properties, and both are abstract. It’s a costly approach, in my view. So there seems to be something wrong with non-self-exemplification, but the source of the problem is elusive.

2. A New Solution

I would like to get a new solution on the table by showing that no contradiction or obvious incoherence is implied by the proposition that non-self-exemplification exists—either as a property or as a meaningful predicate. I will begin by stating more explicitly what I take to be the argument against the existence of non-self-exemplification.[6] I will then expose a weak link in the argument.

For ease of presentation, let ‘SE’ designate the property of self-exemplification, and let ‘NSE’ designate the property of *non-self-exemplification*. (I will talk in terms of *properties*, but nominalists may translate what I say in terms of predicates.) We may now state a *reductio* argument against the existence of NSE as follows:

1. Suppose NSE exists.
2. Then: either (i) NSE exemplifies itself, or (ii) NSE does not exemplify itself.
3. Suppose (i) is true: NSE exemplifies itself.
4. Then: NSE exemplifies NSE.
5. Whatever exemplifies NSE does not exemplify SE.
6. Whatever does not exemplify SE does not exemplify itself.
7. Therefore, NSE does not exemplify itself. (4–6)
8. Therefore, (i) is not true. (Because 7 contradicts 3)
9. Suppose (ii) is true: NSE does not exemplify itself.
10. Then: NSE does not exemplify NSE.
11. Whatever does not exemplify NSE exemplifies SE.
12. Whatever exemplifies SE exemplifies itself.
13. Therefore, NSE exemplifies itself. (10–12)
14. Therefore, (ii) is not true. (Because 13 contradicts 9)
15. Therefore, neither (i) nor (ii) is true, which contradicts (2). (8, 14)

[4] I am grateful to [removed] for pressing me with such examples.

[5] I take this brief statement of the approach from [NAME REMOVED].

[6] I say *the* argument because there really is only one argument—the one that purports to show that non-self-exemplification exemplifies itself *iff* it doesn’t.

16. Therefore, it is not true that NSE exists. (1, 2, 15)

I wish to focus on just one premise: (11). Premise (11) states that whatever does not exemplify NSE exemplifies SE. Is that true? It isn't *strictly* logically necessary. Why believe it? One reason one might accept (11)—the only serious reason I see—is that (11) is implied by the following general principle:

17. For any entity *e*, and for any property *p*, if *e* doesn't exemplify ~*p*, then *e* exemplifies *p*.

To illustrate, take the property of being purple. Suppose my pen fails to exemplify its complement, *not being purple*. Then according to (17), my pen thereby exemplifies *being purple*. Similarly, (17) implies that if NSE fails to exemplify NSE, then NSE thereby exemplifies SE, just as (11) asserts.

Must one accept (17)? Someone could perhaps doubt (17) on the grounds that there are no "negative" properties. The thought here is that things fail to exemplify "negative" properties—properties of the form ~*p*—simply because there are no such properties. I won't defend this reply, however, because it seems to me that there are perfectly meaningful negative *predicates*. Moreover, this reply implies that NSE doesn't exist, which is the very thesis I wish to challenge. I am interested, then, seeing whether someone could have another reason for not accepting (17)—one that doesn't itself imply the conclusion of the argument against NSE.

Someone could refrain from accepting (17) simply because they see no clear reason to do so. Is there a reason to accept (17) that ought to compel everyone who reflects on the matter? Is there even an argument for (17) that should compel *anyone*? Answers are far from obvious.

To help us think about (17), consider this alternative principle:

(Lacks) For any entity *e*, and for any property *p*, if *e* doesn't exemplify *p*, then *e* lacks *p*.

Suppose (Lacks) is true. Then we can make sense of the metaphysics of *not having properties* without invoking the relation of exemplification. So, for example, if one says "Natasha is not hungry," one expresses the proposition that Natasha *lacks* hunger. And this proposition is importantly different from the more complicated proposition that Natasha *has* (exemplifies) *non-hunger*. If (Lacks) is true, then non-hunger may be a coherent concept (and a genuine property) even if non-hunger is not the right kind of thing to be *exemplified*. It could be that I lack hunger without thereby also *having* non-hunger.

Ordinary language suggests that there are two importantly different ways to characterize an object: we may characterize the way an object *is*; or we may characterize the way an object *isn't*. One might think these are fundamentally different modes of characterization, for they involve fundamentally opposed characterization relations (or ties). One such relation (or tie) is exemplification. The other, one might think, is *reverse*-exemplification—also known as *lacks*.

Perhaps, then, certain properties (such as "negative" ones) are *lacked* rather than *had*. If that's right, then it could be that I lack (say) hunger while also lacking—i.e. not *having*—non-hunger.

I should emphasize that the above story can also make sense of our talk of certain predicates failing to be satisfied. For example, Natasha might say "I am not hungry" without thereby committing herself to the further semantic thesis that she satisfies the predicate "is not hungry". Maybe she fails to satisfy the predicate "is hungry", while also failing to satisfy the complement of that predicate. That's certainly conceptually possible. Moreover, we must say something like this with respect to certain predicates. For example, "non-self-describing" fails to satisfy "non-self-describing", while also failing to satisfy its complement. If that were not so, then "non-self-describing" would satisfies itself *iff* it doesn't, which is absurd. So, it makes good sense to allow a predicate and its complement to fail to be satisfied (exemplified or had).

At this point, someone might wonder whether my proposal is analogous to the "truth-value gap" solution to the Liar paradox—which is the theory that Liar sentences, like 'this sentence is false', lack a truth-value. There is this connection: my proposal that a thing may lack a property and its complement empowers the truth-value gap solution, since it enables the proposal that a Liar sentence lacks both truth and the complement of truth. On the other hand, the "truth-value gap" solution famously leads to a "Revenge" version of the Liar paradox: for example, "this sentence lacks truth" is no less paradoxical if we suppose that it lacks a truth value.[7] Fortunately, there doesn't seem to be an analogous "Revenge" version of the problem of non-self-exemplification. The closest parallel I have seen arises from considering the property of *lacking oneself*. We might wonder whether that property exemplifies itself *iff* it lacks itself. But it isn't clear why we should think so, unless we accept the critical premise (17). For if (17) is false, as I have proposed, then *lacking itself* may lack itself without thereby exemplifying its complement, which is *not lacking itself* (or *exemplifying itself*).[8] So by allowing that a property, such as non-self-exemplification, may lack itself and its complement, the paradox falls by the wayside.[9]

The suggestion, then, is that NSE can exist without being exemplified by anything, not even by NSE. NSE may lack SE, I suggest, while also lacking NSE, just as Natasha may lack hunger, while also lacking *lacking hunger*. I am not aware of any compelling argument against this proposal, and therefore I offer it as an option worthy of serious consideration.
I do not claim that every philosopher will be able to accept my proposal. Some philosophers are committed to sparser ontologies, such as ontologies in which there are no properties, or in which there are no properties as unnatural as non-self-exemplification. Other philosophers may balk at

[7] But see [removed] for a recent way to develop the "truth-value gap" solution in a way that escapes paradox.

[8] Note that anyone who adopts this proposal must deny that a property *p* that lacks itself thereby *has* the property *lacking itself*. In general, a property *p* may lack a property *q* without thereby *having* the lacking of *q*. If this were not so, then we would fall into contradiction, since we could then infer by contraposition that *lacking itself* does not lack itself from the opposite proposal that *lacking itself* lacks—and so does not *have*—the property of *lacking itself*. I owe this clarification in response to a query posed to me by [removed].

[9] I am grateful to [removed] for drawing my attention to this potential worry.

the proposal that things do not *have* negative properties. It is an unusual view. On the other hand, the view I propose strikes me as the more satisfying and less ad-hoc than any of the current solutions on the market. I suspect I will not be the only one. A central value of my proposal is that it brings to light a new way to solve the problem of non-self-exemplification, while avoiding significant drawbacks of existing solutions. The proposal also provides a metaphysical foundation for a certain theory of *sets*, as we shall see in the next section.

4. Reflection

Previous solutions to the problem of non-self-exemplification purport to give a principled explanation as to why there is no such property as non-self-exemplification. I have argued instead that there is actually no incoherence in supposing that non-self-exemplification exists.

I would like to close by showing, briefly, how my account of having and lacking properties can be used to motivate a "positive" theory of set formation. We learned from Russell (and others) that the completely unrestricted naïve set theory is too naïve to be true. The completely unrestricted theory says that for any well-defined predicate, there is a corresponding set of things characterized by that predicate. Yet as Russell famously pointed out, the set, {*x* : *x* is not a member of itself}, is a member of itself *iff* it isn't. To avoid the contradiction, we are forced, then, to tighten the conditions of set formation. One relatively simply way to tighten the conditions of set formation is in terms of "positive" set theory.[10] We may suppose, for instance, that for any well-defined *positive* predicate, there is a corresponding set of things characterized by that predicate. So, for example, there is the set of all dogs. But there isn't a set of all things that are *not* a dog, unless there is a positive predicate that applies to all and only those things that aren't a dog.

Positive set theory enjoys the virtue of being both simple and powerful. It is more powerful (and may be closer to our naïve conception of sets) than contemporary theories of sets, such as ZF, ZFC, NF and NFU. Positive set theory replaces ZF's weak separation axiom,[11] for example, with an axiom Esser calls 'Generalized Positive Comprehension':

(GPC) For any generalized *positive* formula φ, $\{x \mid \varphi\}$ exists.

We may treat a positive formula as any formula that has no negative predicates.[12]

---

[10] For a relatively recent example of a "positive" set theory, called 'GPK^{+}_{∞}', see Esser 1999.

[11] Separation: for any formula φ in the language of set theory, and for and any set *A*, there is a set, $\{x \in A \mid \varphi(x)\}$.

[12] Or we could go with this standard account: a 'positive' formula is a formula belonging to the smallest class of formulas that contains (i) a false statement, (ii) atomic membership and equality formulas, and (iii) all conjunctions, disjunctions, existential and universal quantifications formed from formulas in (ii). See Holmes 2012.

So by recognizing a distinction between *having* and *lacking* properties, we gain ontological resources to motivate a powerful theory of sets. We also gain a new solution to an outstanding paradox curled in the heart of human reason.

References

Esser, Olivier. 1999. “On the Consistency of a Positive Theory,” *Mathematical Logic Quarterly*, 45 (1): 105–116.

Holmes, Randall. 2012. “Alternative Axiomatic Set Theories,” *Stanford Encyclopedia of Philosophy*.