

2022 Annual Undergraduate Philosophy Conference Journal

Conference Organizers: Lin Ye, Lucie Lu, Tyler Olds, Jonathan Tanaka, Ryan Wang,
Heyan Xiao, Pacy Yan

Review Committee: Christian Baez, Michael Davis, Sophia Lin, Lucie Lu, Ryan Wang,
Pacy Yan, Kingsley Zheng, Yuanji Zhu

Special thanks to the NYU Department of Philosophy

nyuphilforum.org

nyuaupc.org



Order of Papers

1/ To Hell with the PII: An Argument from Indiscernible Quantum Objects
By Tessa Ng (University of Toronto)

2/ Dissolving Ontic Vagueness: Evans Reconsidered
By Kaidi Pan (University of Chicago)

3/ Duties of Friendship: A Refutation of Shiffrin's Transcendental Argument
For the Power to Promise
By David Sutherland (University of California, Los Angeles)

4/ Belief Polarization: Is Path-Dependent Evidence Search Irrational?
By Shuyuan Wang (University of Chicago)

To Hell with the PII: An Argument from Indiscernible Quantum Objects

By Tessa Ng

University of Toronto

To Hell with the PII: An Argument from Indiscernible Quantum Objects

ABSTRACT

In the metaphysics of identity, the Principle of the Identity of Indiscernibles (PII) is a central doctrine which states that no two things are exactly alike. Introduced by Gottfried Wilhelm Leibniz, the PII serves as a cornerstone to our understanding of the universe and the types of objects that can exist in it. In 1952, Max Black famously declared that he had conjured up a hypothetical universe that refutes the PII and in 1995, John O'Leary Hawthorne responded with a paper of his own, detailing how the Bundle Theory of Substance nullifies Black's revolutionary argument and saves the PII. In this paper, I will outline the contentions of both metaphysicians, refute Hawthorne's argument on the grounds of logical inconsistency, and propose that the strongest blow we can land against the PII stems not from the hypothetical universe proposed by Black, but from small pockets of reality that exist within our universe: the realm of quantum objects.

To Hell with the PII: An Argument from Indiscernible Quantum Objects

INTRODUCTION

Two indiscernible spheres exist at a distance from each other in an otherwise empty universe. Castor and Pollux, the lone inhabitants of the hypothetical world in Max Black's famed thought experiment, are qualitatively and relationally identical. Every property that Castor instantiates is mirrored by Pollux, such that nothing can be said about one that cannot also be said about the other. This is the case presented in Black's 1952 paper, "The Identity of Indiscernibles". The Principle of the Identity of Indiscernibles (PII) states that if two objects x and y possess all the same properties, then x and y are identical. Accordingly, the PII would have it that Castor and Pollux are the same object. However, this raises a contradiction. If Black's universe consists of two distinct spheres and nothing else, how can it be that there is just one object? It seems the case that Castor, Pollux, and Max Black have defeated the PII.

In his 1995 paper, "The bundle theory of substance and the identity of indiscernibles", John O'Leary Hawthorne asserts that this is not the case. Hawthorne argues that Castor and Pollux pose no threat to bundle theory and that employing the bundle theory of substance allows us to save the PII. In this paper, I will contend that Hawthorne's argument from bundle theory is a weak defense of the PII because it contains a logical contradiction and a small amount of circular reasoning. In addition to this, I will propose that quantum objects make trouble for the PII and that we can, at best, defend a diluted version of it for the purpose of parsimony. I will accomplish this by reconstructing Hawthorne's argument, highlighting chinks in the armour of bundle theory, and outlining how the indiscernibility of quantum objects undermines the PII in the real world.

EXEGETIS

At the crux of Hawthorne's argument is the idea that, assuming bundle theory, Max Black's world does not disprove the PII. The bundle theory of substance states that the substances that make up the world are nothing beyond bundles of properties called universals. These universals are qualities that objects may instantiate, excluding haecceities. Examples of universals include the qualities of blueness, squareness, lumpiness, and other characteristics of this sort that multiple objects may instantiate. It is of primary importance to Hawthorne's argument that the bundle theory employs an immanent conception of universals, which asserts that "universals are actually in space and time"¹. In other words, there are substances that occupy space and time and these substances have universals, or qualities, that are also in space and time. A consequence of this conception is that universals can be instantiated by multiple objects and be present in multiple places simultaneously. This is not hard to believe, as it is easy to imagine two objects, such as two adjacent frogs, who mutually instantiate the universal "squatness". Hawthorne expands this line of reasoning to arrive at the claim that, just as a single universal can be instantiated by two distant objects, so can a bundle of universals. That is, the exact same bundle can be distantly instantiated.

Despite its simplicity, the previous conclusion has great consequences for Hawthorne's argument. If one agrees that the same bundle of universals can be distantly instantiated, they ought to admit that the existence of Castor and Pollux is fully viable, given the bundle theory. Moreover, if we accept that bundle theory is true, then the PII is true, and there is a single bundle of universals at a distance from itself in Max Black's universe. Black's thought experiment merely demonstrates that it is possible for one object to exist in two places. Hence, bundle theory saves the PII from Castor and Pollux.

¹ (Hawthorne 1995, 191)

To recap, Hawthorne's argument can be expressed in premise-conclusion form as follows:

- (1) The PII states that if two objects x and y possess all the same properties, then x and y are identical.
- (2) The bundle theory states that the substances that make up the world are bundles of universals.
- (3) The bundle theory deploys an immanent conception of universals, which states that the same bundle of universals can be instantiated in multiple places simultaneously.
- (4) Castor and Pollux instantiate all the same properties and are at a distance from each other.
- (5) CONCLUSION: Given bundle theory, the PII is true and Castor and Pollux are the same bundle of properties at a distance from itself.

CRITICISM FOR HAWTHORNE'S ARGUMENT

At this point in the paper I will discuss weaknesses in Hawthorne's argument. Firstly, I reject Hawthorne's third premise (3), which entails that one substance can exist at a distance from itself. Although it is conceivable that two identical but distinct, or two *very* similar substances, may occupy different spatial locations, the notion of a single substance occupying multiple non-equal locations simultaneously seems inherently contradictory. Distance, as a relation, implies distinctness, for how could a single object (that is not a wave or a field) be spatially separate from itself²? The contradictory nature of being at a distance from oneself can be illustrated through the following example: Consider the logical tautology "I am here". Say the speaker is a single object that occupies two distinct locations a and b such that we refer to a as "here". The point b is then "not here" because $a \neq b$. To describe their location, the speaker must say "I am here" *and* "I am not here", which is a logical contradiction. Hence, I reject Hawthorne's third premise (3) for logical inconsistency.

The second bone I pick with Hawthorne's argument is that he uses the PII to save the PII in a question-begging loop of flagrant circularity. There is an indecorous logical gap between two claims that

² In fact, there are ways this happens, such as quantum superposition, but I would argue that such a phenomenon only has sway at the quantum level and will not bear on objects such as Castor and Pollux.

are made at the bottom of page 193. In the interest of concision, I have reconstructed the argument as a series of claims:

Claim 1: The same universal can be multiply instantiated by two objects at a distance.

Claim 2: Bundles of the same universals can be multiply instantiated at a distance.

Claim 3: The same single bundle of universals can be instantiated at a distance from itself.

Notice here that the PII is a necessary implicit claim needed to get from Claim 2 to Claim 3. In Claim 2, we have that it is possible for there to be two identical bundles at a distance, but it has not yet been argued that there is only one bundle present. In Claim 3, Hawthorne makes a statement about a single bundle. In order to get from Claim 2 to Claim 3, Hawthorne implicitly assumes the PII (as a sort of Claim 2.5) and reasons that, given PII, two bundles of the same universals are a single bundle. He then claims that this does not contradict the PII and that bundle theory has, hence, saved the PII from Max Black. Of course a claim arrived at using the PII will not contradict the PII, this is circularity. Hence, I reject Hawthorne's argument that the bundle theory of substance saves the PII from Castor and Pollux because his argument is circular.

To Hell with The PII PII: AN ARGUMENT FROM QUANTUM OBJECTS

One objection that has been raised in opposition to Black's argument against the PII is that Castor and Pollux are a mere thought experiment and that such a world does not really exist, so Black's conclusion should not be taken seriously. Consider now a new Max-Blackesque world, except this one exists somewhere inside our own: the world that is a snapshot in time of a single helium atom. Two identical electrons exist in the probability cloud of a helium atom, call them Pastor and Collux. In every qualitative respect, the electrons are identical. Pastor and Collux have the same charge, mass, spin, and are entirely indiscernible. Moreover there is a unique symmetry to this snapshot of a helium atom, such that it was taken when Pastor and Collux were oppositely equidistant from a perfectly symmetrical

nucleus. In other words, if you “sliced” the atom down the middle, each half would be a mirror image of the other and contain 1 proton, 1 neutron, and 1 electron, with all relative distances equal to the other half³. With nothing else in this world to distinguish Pastor and Collux, in light of Hawthorne’s argument, we can predict that the bundle theorist would say there is a single electron in this system. However, this certainly cannot be the case because many of helium’s chemical properties are a result of the specific number of electrons it has. In order for this to be a helium atom, as it is, there necessarily must be exactly two electrons in the system. Do Pastor and Collux successfully constitute a real-world counterexample to the PII? Unfortunately, not. Due to the Pauli exclusion principle, we know that the two electrons occupy different quantum states as any two fermions in a system must and are, hence, discernible by this quality. However, Pastor and Collux are a useful tool for illustrating the line of reasoning I intend to take to form an argument against the PII from quantum objects.

Consider now the universe that contains a two-photon “beam of light” and is otherwise empty, this is similar to considering the isolated system of light in a vacuum. Of course, this beam will not meet the standards of what is commonly considered an adequate light source, but no matter. What matters is that there are two photons and nothing else in the system we are considering. As an homage to Black, call the two photons Romulus and Remus. The photons are once again indiscernible objects: they are both massless, have spin-1, the same energy, and are moving along the same trajectory at the same frequency. The system is entirely symmetrical. Moreover, since photons are bosons, they are able to occupy the same quantum state, so Romulus and Remus, in fact, do occupy the same quantum state. These two photons are what are known as “indiscernible particles”. Yet again, the bundle theorist wants to say that there is only one photon here and that it propagates through space beside itself. However, if the beam of light were to be measured, we would find that the intensity of light is twice what it would be if there were only one photon in the beam. Knowing that the intensity of light is directly proportional to the number of photons

³ This is quite a fantastical thought experiment, though considering the probabilistic nature of the states of atoms, it is believable that such a system has existed for an instant in our world at some point.

in a beam, we know that there really are two photons present ⁴. It seems we have conjured up a scenario that acts as a counterexample to the PII.

Interestingly, many similar cases of quantum objects undermining the PII can, and do, take place in the real world. These scenarios are not without counterarguments. For instance, an interesting explanation of the existence of indiscernible electrons was posited by American physicist John Wheeler in 1940. Wheeler's One Electron Universe theory postulated that every electron in the universe is actually the same single electron tracing out a knot-like oscillatory world line through space and time⁵. Counterarguments, such as Wheelers, are of minimal concern because the theories that support them rests on far more precarious physics than the well-tenured science that supports the helium atom example, as well as the two-photon beam example. At best, counterarguments to cases of quantum objects that undermine the PII attack spatiotemporal relations, which are fuzzy, incoherent, and highly probabilistic at the quantum level. Although it is of metaphysical interest to defend the PII for its parsimonious consequences, quantum indiscernibles pose a formidable foe to proponents of the principle.

CONCLUSION

Hawthorne has argued that the bundle theory of substance is able to save the PII from Castor and Pollux of Max Black's famed thought experiment. Although Hawthorne's argument did an excellent job of dealing with the case at hand and was appreciably parsimonious, it ultimately provides a feeble defense of the PII due to an internal contradiction and circularity. In the tradition of Max Black, I propose that there exist, indeed, many counterexamples that undermine the PII and that some interesting instances of these can be found by looking at the indiscernibility of quantum objects.

⁴ This thought experiment is very "hand-wavy" and idealized, but sufficiently plausible for the sake of postulation.

⁵ (Feynman 1995)

BIBLIOGRAPHY

O’Leary-Hawthorne, John. “The Bundle Theory of Substance and the Identity of Indiscernibles.”

Analysis 55, no. 3 (1995): 191–96. <https://doi.org/10.2307/3328579>.

Black, Max. “The Identity of Indiscernibles.” *Mind* 61, no. 242 (1952): 153–64.

<http://www.jstor.org/stable/2252291>.

Feynman, Richard. “Nobel Lecture”. (1995).

<https://www.nobelprize.org/prizes/physics/1965/feynman/lecture/>

Giglio, Daniel Joseph. “How to Think About Indiscernible Particles.” (2013). *Theses and Dissertations*.

103. <https://dc.uwm.edu/etd/103>

Dissolving Ontic Vagueness: Evans Reconsidered

By Kaidi Pan

University of Chicago

Abstract:

Evans (1978) offers a formal argument suggesting that *de re* indeterminate identity statements are impossible. Lewis (1988) offers a reply to Evans' argument. This paper will build on Evans' argument and Lewis' reply and argue that Evans' argument – if correctly understood – does succeed in rendering ontic vagueness in vain. The idea of ontic vagueness leads to incoherence; other precisifications of ontic vagueness either are practically unapproachable or can be equally well-modeled as a linguistic phenomenon.

Dissolving Ontic Vagueness: Evans Reconsidered

I The Debate

Vagueness is a ubiquitous phenomenon in human language. One distinctive feature of vague predicates is the susceptibility to borderline cases in their extension, consider the sentence:

(0) x is bald

There exists some x such that it seems equally plausible to assign (0) a truth value of either T or F. Attempts to model this phenomenon has been made, including the three-valued logic, supervaluationism, contextualism, etc. Numerous such attempts share a similar insight to the nature and origin of this phenomenon – vague predicates are only because of the fuzziness in our representation. Representational vagueness can be defined as such:

Def. 1.1 We say vagueness is representational, if and only if, for any sentences S involving predicate P, if S is vague, then S's vagueness is the result of indeterminacy in how we represent the world¹.

We can see how this definition applies to the prevailing semantic model of vagueness. Supervaluationism, for example, suggests vague predicates are the result of incompleteness in their meaning and subject to precisifications; were precisifications coherent with respect to the established semantic decisions, vagueness could be resolved². This shows that for supervaluationists, vagueness is a result of semantic indeterminacy, which implies it is

¹ Some may argue that there are different kinds of representational contents. In this paper, we are concerning semantics as the primary representational contents. For a similar definition, see Sainsbury, (1995)

² See Fine, (1975)

representational.

On the contrary, given the same phenomenon of vagueness, proponents of ontic vagueness suggest that the vagueness in the predicates are not representational, i.e., not because of the incompleteness/indecision/fuzziness implicitly in the semantical content of the predicate; there are vague predicates because of ontic vagueness, i.e., the object that the predicate describes is indeterminate *de re*.

Def. 1.2 We say vagueness is ontic, if and only if, for any sentence S involving predicate P, if S is vague, then S's vagueness is the result of indeterminacy of how the object that P described is P-wise³.

It is important to note here that vagueness is primarily a phenomenon of our language. No one would deny that vagueness is *prima facie* a feature of the predicates and thus the sentences involving that predicate. Therefore, both definitions are *not* defining where vagueness happens, but defining where the fundamental sources of vagueness are, i.e., whether vagueness in our language is ontically-induced or representationally-induce. However, def. 1.2 is not entirely clear: when we say an object is *P-wise*, we could still mean a) this object is vaguely *P* because the word *P* is vague, b) this object is vaguely *P* because the object is vague. Ontic vagueness, which suggests that the source of vagueness in our languages is the indeterminacy of the object that the predicate is *precisely*⁴ fixed onto, which intends to mean b. With this discussion in mind, we can have a more substantial definition:

Def. 1.2.1 We say it is indeterminate that the object that P described is P-wise, if and

³ For a similar definition, see Barnes, (2010)

⁴ We will discuss this “*precision*” in further detail, i.e., we want the reference be so determinant such that it is a relation of rigid designation. See section IV.

only if, when all representational content is precisified and there exists an admissible precisification of P , according to this precisification, the sentence including P is still indeterminate⁵.

Arguably, def. 1.2.1 successfully captures the sense of ontic vagueness, i.e., we cannot locate ontic vagueness until we are sure that our language is precise enough to rule out representational vagueness.

Now that we have set up the debate: whether vagueness is representational or ontic. Evans' article *Can There Be Vague Objects* purports to be a proof of the impossibility of *de re* indeterminate identity statements⁶. In this paper, we will expose two interpretations of Evans' argument in favor of the representational theorists; we will also anticipate a possible response on behalf of ontic theorists. Overall, we argue that (following Evans' argument) ontic vagueness is either a guise of representational vagueness or an incoherent idea in itself; otherwise, we face a practical difficulty to locate ontic vagueness.

II Evans' Argument

In his article *Can There Be Vague Objects*, Evans proposes the following argument⁷:

We define a sentential operator ∇ which stands for "it is indeterminate that..."

$$(1) \nabla(a = b)$$

it is indeterminate that a is identical to b

$$(2) \lambda x[\nabla(x = a)]b$$

⁵ For a similar definition, see Barnes, (2010)

⁶ See Evans, (1978), and Barnes, (2008)

⁷ See Evans, (1978)

b is such that being indeterminately identical to a

$$(3) \neg \nabla(a = a)$$

it is not indeterminate that a is identical to a

$$(4) \neg \lambda x[\nabla(x = a)]a$$

a is not such that being indeterminately identical to a

$$(5) \neg(a = b)$$

It is not that a is identical to b

The conclusion (5) is achieved by combining (2) and (4) with Leibniz's Law, which suggests the following: if $a = b$, then all the properties of a 's are the same as all the properties of b 's. From two, we have: b has the property of being indeterminately identical to a , whereas a does not have the property of being indeterminately identical to a ; therefore, a and b do not share all of their properties, and thus it is not the case that a is identical to b by the contraposition of the Leibniz's Law.

The aim of this line is supposed to question if the view "the world might contain certain objects about which it is a fact that they have fuzzy boundaries" is coherent; therefore, the reason why Evans starts with (1), i.e., supposing a vague identity statement with the indeterminacy operator, should be understood as to a supposition on the side of ontic vagueness, and a proceed the proof by contradiction.

Two clarificatory questions should be answered before proceeding to examine whether this is a good argument. First, how exactly (5) can be viewed as a claim that leads to a contradiction? Evans suggests that we define the dual of ∇ , Δ which stands for "it is determinate that..."; and let the operator ∇ , Δ generates an S5 axiomatic system; that is,

different from the S5 in modal logic, we replace the contingency operator with ∇ , and the necessary operator with Δ , we call this system S5* for notational convenience (because we will use the S5 system for both operator pairs later in this passage). Then, (1), (3) can add Δ in the front by axiom 5 of S5*⁸, and hence (2), (4) is determinate. Therefore, then we can derive the following (5*):

$$(5^*) \Delta \neg(a = b)$$

It is determinate that it is not that a is identical to b

(5*) is a contradiction to (1). Yet another question remains – what does this contradiction achieve? (1) asserts vague identity statements, and Evans' argument has just shown that statements in (1)'s form leads to a contradiction. Therefore, there can be no vague identity statements. A missing step in Evans' argument is to explain how the impossibility of a vague identity statement can be viewed as a challenge to the general idea of ontic vagueness.

I believe the following lines should complete the argument: suppose vagueness is ontic, then, by def 1.2, we have a vague object X such that it is indeterminately P -wise. Consider an object X' such that it is also indeterminately P -wise *ceteris paribus*, which lies somewhere at the borderline just like X . By def 1.2.1, we want to say P is precise, and the sources of indeterminacy are in object X and X' . Then, it is indeterminate whether X fully coincides with X' , which implies (1) it is indeterminate whether $X = X'$.

Consider the following scenario:

Suppose there is a cloud, and a cloud is a vague object because there are a small number of molecules on the borderline of the cloud; and assume it is equally plausible to say that the

⁸ By replacing the contingency operator with ∇ , and the necessity operator with Δ , axiom 5 of S5* can be expressed as: $\nabla A \rightarrow \Delta \nabla A$, which can then be applied to (1) and (3)

cloud includes or excludes these molecules⁹. We name the cloud excluding those borderline molecules C1 and the one including these molecules C2. Since it is indeterminate whether the cloud contains these molecules, it is indeterminate whether C1 is identical to C2.

The brief argument above demonstrates why if there is ontic vagueness, we have vague identity statements. Therefore, the overall idea is that if we think there exist vague objects, we must have a vague identity; and if Evans' argument works, then vague identity cannot be the case, which by contraposition refutes the idea that vagueness is ontic.

In the next two sections, I shall evaluate Evans's argument based on two interpretations, both concerning the inferential step from (1)-(2) and from (3)-(4). From there, we can see how both interpretations favor the representational side of the debate.

III Reconsidering Evans: Option 1

Option 1: we should interpret the a or b in Evans' argument to be non-rigid designators¹⁰, which means, a or b refer to different objects across the different possible world, and under different precisifications.

Under this interpretation, Evans's argument is invalid because of the following problem:

In S5*, as we have discussed in section I, by applying axiom 5 to (1) we have:

$$(1^*) \Delta \nabla (a = b),$$

The supposed (2*) inferred from (1*), if Evans's argument works, i.e., successfully derive a contradiction, should be:

$$(2^*) \Delta \lambda x [\nabla (x = a)] b$$

⁹ See Lewis, (1999)

¹⁰ See Lewis, (1988)

However, if we understand *a* and *b* to be non-rigid designators, the inferential step from (1*) to (2*) is fallacious. In (2*), *b* moves outside of the scope of the indeterminacy operator, which is wrong.

(1*) suggests it is determinate that it is indeterminate that *a* is identical to *b*; whereas (2*) suggests it is determinate that *b* has the property of being indeterminately identical to *a*. There are some substantial examples indicating the impossibility of this move, for better semantic intuition, let us consider these two sentences in S5 system (since S5 and S5* are formally the same, the invalidity of (1)-(2) is applicable to both):

(1*') (It is necessary that) it is contingent that the height of Mount Everest is 8848m. (True)

(2*') (It is necessary that) the height of Mount Everest is such that being contingently 8848m. (False)

In this example, I replace *b* to be a non-rigid designator (let the one moving outside of the scope of indeterminacy operator from (1) to (2) to be non-rigid), and we can see the problem: (1*') and (2*') does not have the same truth value; (1*') is a *de dicto* claim and (2*') is a *de re* claim. The height of Mount Everest is not a rigid designator strictly referring to 8848m in all possible worlds (parallelly in S5*, under all precisifications), and thus it is contingent – it only suggests that the height of Mount Everest contingently refers to something, i.e., 8848m; but that does not in any way implies that the height of Mount Everest is necessary being contingently 8848m – which suggests that the height of Mount Everest necessarily refers to a thing, i.e., the thing of “indeterminately 8848m”.

Therefore, parallelly in S5*, the height of Mount Everest is not referring to 8848m under all precisifications, which makes it indeterminate; however, it does not imply that the height of

Mount Everest is determinately being indeterminately 8848m. In other words, the sentence “the height of Mount Everest indeterminately refers to something, i.e., 8848m”, is different from the sentence “the height of Mount Everest determinately refers to something, i.e., indeterminately 8848m”. Consequently, it seems like Evans’ argument is wrong under this interpretation, i.e., viewing a or b to be non-rigid designators.

While it is true that Evans’ argument is invalid under this interpretation, it does not effectively undermine Evans’ goal, because if we view a or b to be non-rigid designators, we fall back to appeal to the more traditional representational view of vagueness. Consider the following sentence from the cloud example in section II:

(6) It is indeterminate that the cloud is identical to C1

The cloud is non-rigid, meaning that it refers to different objects across different precisifications, i.e., C1, C2, or etc. However, if this is the case, then the following move can be made on behalf of representational theorists:

If the cloud is non-rigid, then it refers to different objects under different precisifications. For example, under some precisification the cloud is C1, and under some precisification, the cloud is C2 (the one including all the borderline molecules); but then this scenario is exactly what supervaluationists consider to be neither super-true nor super-false: under some precisifications the cloud is C1, and under some precisifications the cloud is C2, but there exists no admissible precisification such that can make either of them super-true. This already is a sufficient model for the entire scenario: the reason why we utter the vague identity statement (6) is precisely because at least one name in (6), i.e., the cloud, is non-rigid.

Yet the non-rigidity of this name *eo ipso* constitutes an account of vagueness that can be

modeled via a representationalist framework, i.e., the representationalist can regard this non-rigidity as a referential indeterminacy rather than ontic indeterminacy, and referential indeterminacy can be viewed as the incompleteness of the meaning as well. For example, by modeling the vagueness of “the cloud” by the supervaluationist method above, non-rigidity itself is considered to be the source of vagueness; non-rigidity of this statement is essentially considered as an incompleteness of its meaning such that it is precisification-susceptible, e.g., making precisifications to “the cloud” is to complete the meaning of it, making it rigid.

Even if the statement ends up in the truth-value gap after precisifications such that it is neither true nor false, i.e., “the cloud” is indeed vague, it still suffices to support the idea that vagueness lies primarily in our representation of the world because we can understand it to be semantic indecision¹¹. This is because being non-rigid implicitly is a semantic phenomenon, and therefore it is susceptible to representationalist approaches such as supervaluationism. As a result, the susceptibility of representationalist approaches supports the representationalist view, i.e., vagueness is not an ontic phenomenon but something in our modes of representation (by def. 1.1).

The idea is that if a or b is non-rigid, then Evans’ argument fails because the move from (1)-(2) is not truth-preserving; but this fact does not immediately constitute support for ontic theorists on the contrary. Because whenever we admit a or b is non-rigid in a sense that under all precisifications it refers differently, then representational theorists can thereby claim this non-rigidity to be certain referential indeterminacy rather than ontic indeterminacy that is susceptible to precisifications, which still supports the idea that vagueness is representational.

¹¹ See Fine, (1975)

IV Option 2

Option 2: we should interpret the a and b in Evans' argument to be two rigid designators, which means, a and b directly refers to objects across all possible worlds (in S5), and under all precisifications (in S5*).

The motivation behind this interpretation is that the inferential step from (1) to (2) requires us to move a and b outside of the scope of the indeterminacy operator. Rigid designators, unlike the non-rigid ones, can move outside of the scope of the indeterminacy operator; and therefore, had a and b both be rigid designators,¹² the inferential step from (1)-(2) could have been made. The following argument through semantics¹³ should justify this move:

If a and b are rigid designators to object α and β , then, in S5* a and b determinately refers to them, i.e., under all precisification a refers to the same object α , and b refers to the same object β . Therefore, when we say (1*) $\Delta\nabla(a = b)$, we mean: it is determinate that it is indeterminate that α is identical to β - a *de re* claim. When we say (2*) $\Delta\lambda x[\nabla(x = a)]b$, we mean: it is determinate that β is such that being indeterminately identical to α .

In option 1, we say (1*) we mean something *de dicto* such that: it is determinate that is indeterminate that a is identical to what b refers to because the non-rigidity of b allows the room for referential indeterminacy, i.e., we are not sure whether b determinately refers to β . It is strictly because of this referential indeterminacy we are not allowed to move this non-rigid

¹² It should be noted here: unlike option 1 which only requires one of the a or b to be non-rigid, option 2 requires a and b to both be rigid designators.

¹³ The reason why here this argument needs semantic interpretations to proceed is that there are no consensual grammatical rules that both involves S5 and λ -calculus. Therefore, the best we can do for now is to proceed under an interpretation.

designator out of the scope of the indeterminacy operator, because in (1*), the referential indeterminacy of the non-rigid designator b is preserved, yet in (2*), referential determinacy disappear.

Option 2 on the other hand suggests, by asserting rigid designation, we *eo ipso* eliminate the possibility of referential indeterminacy of option 1's kind. The only reason why we cannot move b outside the scope of ∇ is that b is indeterminately referring to something in (1*), but determinately referring β in (2*); therefore, as we assert the rigid designation, i.e., b is indeed determinately referring β in (1*) just like (2*), the moving of b out of the scope of ∇ is truth-preserving because under this reading Evans' argument resolutely excludes all referential indeterminacy before the inferential step from (1)-(2) (or (1*) to (2*)).

But what exactly α and β are supposed to be? As we have argued, if we have (1) $\nabla (a = b)$ where a and b rigidly designate α and β , the only way we can make sense of this statement is "it is indeterminate whether α is β " – whenever we make sense of this statement, we always pointing at ontic vagueness such that α and β are indeterminately identical because they are vague objects, *not* that a and b are vague names. Therefore, the vague identity statements such as (1) and (1*), under this reading, imply the thought that we have vague objects – a and b are rigid designators, designating vague objects.

Consider the following example for a better semantic intuition (Sagarmatha is Mount Everest in Sanskrit):

(1*") (It is determinate that) it is indeterminate that Mount Everest is Sagarmatha (True)

(2*") (It is determinate that) Mount Everest is such that being indeterminately Sagarmatha (True)

We can see that if (1*) is true, (2*) is true. We know Mount Everest rigidly designates something, and Sagarmatha rigidly designates something; therefore, were (1*) is true such that it is indeterminate that Mount Everest is Sagarmatha, the indeterminacy is not because 1) Mount Everest indeterminately refers to something, or 2) Sagarmatha indeterminately refers to something, but because Mount Everest and Sagarmatha designate something that is in itself indeterminate, i.e., a vague object. (2*) suggests the vague object Mount Everest designates is such that it is indeterminately identical to the vague object Sagarmatha designates – this is also true for the reasons we discussed above.

There is immediate merit of this interpretation. As we read a and b transparently such that a and b rigidly designate objects across various precisifications, the vague identity statement (1) and (1*) immediately boils down to ontic vagueness: by def. 1.2 and def. 1.2.1, we have ontic vagueness - if and only if - despite having a perfectly precisified and admissible language there are still indeterminacies; and under this reading, we do have perfectly precisified and admissible language such that a and b rigidly designate α and β under all precisifications, we still have vague identity statement such as (1) and (1*).

Nonetheless, under this interpretation, given that the inferential step from (1) to (2) and (1*) to (2*) is truth-preserving, the immediate outcome of Evans' argument is that we can derive (5*) $\Delta\neg(a = b)$, which directly contradicts (1*) $\Delta\forall(a = b)$. Therefore, we ultimately find out that, under this interpretation where we did successfully establish an account for ontic vagueness, we arrive at a contradiction; and Evans' argument in section II is successful¹⁴.

The conclusion of this section is thus that if we suggest 1) we should read a and b to be

¹⁴ See Lewis, (1988)

rigid designators, 2) *a* and *b* rigidly designates vague objects, and 3) there is indeed ontically vague identity statements, then inevitably we have to embrace the outcome of Evans' argument in section II: vagueness is not ontic. Therefore, option 2 also seems to support the representationalist side of the debate laid out in section I.

V A Middle Path?

Ontic theorists seem to have a dilemma at this point: on one hand, if they want to point out certain invalidity of Evans argument, they risk of falling to representational vagueness; on the other hand, if they do want to maintain ontically vague identity statement even if all the representational indeterminacy is removed, Evans' argument against them is successful. Either way we find there is something inherently wrong with the idea of ontic vagueness.

However, are the two options we are considering exhaustive? In option 1, we say *a* or *b* is non-rigid, which therefore leads to the idea that we only have the referential indeterminacy – something representational; in option 2, we say *a* and *b* are rigid, which therefore leads to the idea that we only have vague objects – something ontic. But can referential indeterminacy and vague object co-exist such that we both have representational vagueness and ontic vagueness?

This might be possible. One can argue against the following claim:

(7) If *x* is referentially indeterminate, then it is so because of representational vagueness.¹⁵

The thought is, our argument in option 1 seems to rest on (7). We want to say the non-

¹⁵ J. R. G. Williams argues against the idea that “If *a* is referentially indeterminate, then it is so in virtue of semantic indecision”. (7) is a paraphrase of this statement into the terminologies we are using in this paper, see Williams, (2007)

rigidity of x leads to referential indeterminacy, and as referential indeterminacy is a problem of our language, i.e., the way we represent the world, it is, therefore, a scenario of representational vagueness.

If ontic theorists can find a counterexample such that x is referentially indeterminacy in virtue of ontic vagueness, it would suffice to refute (7); and therefore, it seems plausible for them to take a middle path between option 1 and option 2, i.e., we have both representational vagueness and ontic vagueness at the same time, which saves the room for ontic vagueness.

Consider a candidate of such counterexamples by Williams¹⁶: suppose there is an amoeba A such that it undergoes fission, which A splits into two daughter amoebas W and E . Let us assume X , W , and E are completely identical in terms of their shapes, sizes, and other properties; after the fission, W moves west, and E moves east. In this case, we know A survives the past the fission, but:

(8) It is indeterminate that A is identical (continue to survive) to W or E

In this scenario, W and E suffer referential indeterminacy in that W picks out X under one precisification and picks out the new amoeba under the other, and the same indeterminacy applies to E . However, this referential indeterminacy can be understood as ontically-induced: it *seems* like it is *not* that we have imprecise languages, e.g., the incompleteness of the meaning of W or E so that we do not know which one picks out the surviving A (A is the rigid designator of the original amoeba), that causes this referential indeterminacy; more plausibly, it *seems* like we have this referential indeterminacy because of something ontically indeterminate, i.e., whether A survives as W or E .

¹⁶ It should be noted here that we do not aim to criticize Williams' theory of ontic vagueness – we are strictly borrowing the exemplification from him to demonstrate a candidate for refuting (7), see Williams, (2007).

Therefore, this scenario *seems* to achieve the following: we have a case of referential indeterminacy that is ontically-induced. If we just described this amoeba example rightfully, then it suffices to prove (7) wrong. Then it follows that we can have representational vagueness and ontic vagueness together under a single setting, which implies there might exist a middle path between option 1 and option 2 that saves room for ontic vagueness.

Yet we do not consider this scenario compelling - we can still say the referential indeterminacy is representational, it is just that the amoebas are carefully chosen examples that make representational vagueness difficult to discover.

The amoeba's case has two important aspects: 1) it involves an *unknown* event of fission that makes us indeterminate about whether *A* before such change is identical to the either *W* or *E*, and thus, 2) the vague identity statements involve an imprecise assertion of continuity before and after the change, i.e., the survival of *A* into *W* or *E*.

On one hand, I opaquely know there is an event of change that happens, i.e., there exists an event such that *A* splits into *W* and *E* – however, I do not know everything about this fission event, i.e., I do not know what molecules of *A* go to *W*, what molecules of *A* go to *E*, where and what energy did *W* and *E* absorb during the fission, what are the criteria of “surviving”, and etc. On the other hand, I do seem to assert that there is a continuity between *A* and *W* and *E* - when I say it is indeterminate whether the *A* “survives as” *W* or *E*, I implicitly assert there is a continuity between them.

With this outline in mind, we can locate what exactly is the source of referential indeterminacy. It is plausible that the reason why we are indeterminate whether *W* or *E* picks out the surviving *A*, is that we do not know 1) everything about fission, and hence 2)

whether we are right to assert the continuity¹⁷ between *A*, *W*, and *E*. Now we shall offer a possible treatment of the amoeba's case on behalf of representationalists.

Suppose I observe everything about the fission, i.e., I observe what molecules of *A* go to *W*, what molecules of *A* go to *E*, where and what energy did *W* and *E* absorb during the fission, what are the criteria of "surviving", etc. Then, it is plausible to say that we have a standard for the survival of *A*, and we can verify whether the process of fission fits the precisification that *A* survives into *W* or *A* survives into *E*. For example, we can say *A* survives as the one that contains more molecules that belong to *A* during the fission; then, by comparing the number of molecules of *A* that move west to the number of molecules that move east with this standard, we therefore precisified this referential indeterminacy – we were not sure whether it is *W* or *E* that picks out the surviving *A*, but after this precisification, we have a clear answer. What's even more important here is that we are precisifying via a representationalist way, i.e., setting up criteria of survival – which is representational.

Perhaps we can even verify that *A* just no longer exists after the fission, for example, if we hold an extremely rigorous criterium of survival such that for *A* to survive through fission it means 99% of the original molecules are preserved in one of its daughters, then, by comparing to what we know about the fission process, e.g., only 75% of the molecules of *A* goes to *W*, and the rest goes to *E*, then, we can even conclude that *A* failed to survive. This is also a way of precisifying referential indeterminacy, and essentially what we are doing is to set up criteria of survival.

¹⁷ In this part of the discussion, continuity and survival are used interchangeably – for our present purpose they are just the same. We say there is a continuity between *A* and *W* iff *A* survives as *W*; the criteria of survival that we will discuss later is the same as the criteria for continuity, i.e., what must be fulfilled if we say *A* continues as *W* (survives as *W*).

There are plenty of ways in which we can precisify the referential indeterminacy in the amoeba's case, but the upshot of this lengthy discussion is irrelevant of which specific way we choose to precisify: we simply cannot see why we would fail to locate the referential indeterminacy from something representational. The indeterminacy of whether *W* picks out *A* or *E* picks out *A* is strictly an indeterminacy of what criteria we are using to measure *A*'s survival, which can be representational¹⁸.

We are tempted to think that it is ontically indeterminant whether *A* survives as *W*, or *E* because the amoeba's case is framed in such a way that renders on our ignorance of the fission process and the criteria of survival. However, had we taken a step back and asked "Is there any possibility to representationally account for this seemingly ontic indeterminacy," we could have found ways to achieve it, for example, by the treatment I provide above.

Therefore, so far in this section, we have examined the amoeba's case which intends to demonstrate how we can leave room for ontic vagueness even if there exists referential indeterminacy – the strategy is to argue against (7); the amoeba's case does *seem* to show that we can have referential indeterminacy caused by ontic vagueness. However, I argue, by providing a representationalist treatment to the amoeba's case, that it can still be understood and precisified via a representationalist way. In the next few paragraphs, we will try to explore why this "middle path," i.e., we can have referential indeterminacy and ontic vagueness simultaneously, is not a favorable option.

Let us restate the definition of ontic vagueness:

¹⁸ It should be noted again that the treatment we offer above is in no way exhaustive – there are perhaps other ways for us to know which amoeba survives. The point of the treatment above is to demonstrate via exemplification that we can make the *seemingly* ontic indeterminacy representational.

Def. 1.2 We say vagueness is ontic, if and only if, for any sentence S involving predicate P , if S is vague, then S 's vagueness is the result of indeterminacy of how the object that P described is P -wise.

Def. 1.2.1 We say it is indeterminate whether the object that P described is P -wise, if and only if, when all representational content is precisified and there exists an admissible precisification of P , according to this precisification, the sentence including P is still indeterminate.

The concern, as we have discussed in section I, is that we can locate ontic vagueness until we are sure that our language is precise enough to rule out representational vagueness. However, immediately it seems like we have a problem, in this section, we want to explore scenarios such that: we can have referential indeterminacy as well as ontic vagueness, yet this already rules out the possibility of ontic vagueness by definition. Since the only way we can locate ontic vagueness is to rule out any representational vagueness, this option puts us into an apparent unsatisfiable position.

The problem here, we believe, is a *practical* one. It is precisely through a careful discussion of the amoeba's case can we understand why def. 1.2.1 is defined in such a way. We define ontic vagueness with respect to def. 1.2.1 because it intends to solve a practical problem – *not* the problem of whether ontic vagueness is something that really exists in the world, but the problem of how we are able to really locate ontic vagueness. In a situation where we try to save room for ontic vagueness whilst remaining a vague, not-precisified representational content, then we would face the practical difficulty to locate which one of them is the ultimate source of the indeterminacy at hand. In other words, the ontic vagueness asserted in the middle

path is unapproachable and inconsequential for we can model everything representationally.

Our treatment of the amoeba's case precisely elucidates such practical difficulty: for a given referential indeterminacy, ontic theorists can surely argue that ontic vagueness, i.e., the indeterminacy of whether A survives as W or E , is the source of such referential indeterminacy; however, the problem is *not* that ontic theorists are wrong, but that we can find ways in which the same referential indeterminacy can be understood, even resolved, via a representationalist framework, e.g., precisifying with respect to the criteria of survival and etc. Therefore, our aim of presenting treatment to the amoeba's case is *not* to prove that ontic theorists are wrong, i.e., there is no such thing as vagueness in the world, but to present them a practical difficulty that we can always find some representationalist ways to dissolve their alleged ontic vagueness.

It is precisely because of this practical concern that we embrace a definition for ontic vagueness as def. 1.2.1, i.e., we say vagueness is ontic, only if we can locate indeterminacy when there is no vagueness in our representation already. However, through this paper we discover the dilemma implicitly derived from Evans' argument: option 2 of Evans' argument suggests that this view leads to incoherence, and the other options, i.e., option 1 and the middle path, do not even fulfill the definition because the vagueness here is merely representational.

VI Concluding Remarks

We will review what we have achieved throughout this paper:

First of all, we set up the debate between the representationalist and the ontic theorists, we identify the issue at stake is *not* where vagueness exists, but where vagueness comes from

– whether vagueness in our language is representationally-induced or ontically-induced.

We introduced Evans' argument for the impossibility of *de re* indeterminate identity statements, and we saw there are two apparent ways to interpret Evans' argument:

Option 1 suggests that we view the *a* or *b* in Evans' argument as non-rigid designators, and what immediately follows is that Evans' argument is invalid because the move from (1) to (2) and (1*) to (2*) is not truth-preserving. However, the vagueness that occurs in this scenario is only referential indeterminacy which can be modeled via a representationalist framework.

Option 2 suggests that we view the *a* and *b* in Evans' argument as rigid designators, i.e., refers to the same vague object under all precisifications. As this option preserves the inferential validity from (1) to (5) and (1*) to (5*), the conclusion of Evans' argument also favors the representationalist, because *de re* indeterminate identity statements are impossible, and ontic vagueness leads to a contradiction.

We also investigated the possibility of the middle path between option 1 and option 2 such that referential indeterminacy and ontic vagueness co-exist. However, through examining the amoeba's case, we understood that there is a practical difficulty for us to locate ontic vagueness; and we realized that def. 1.2.1 is designed in such a way as to exclude this intractable scenario.

In conclusion, following Evans' argument, we argue that the representationalist side of the debate is more compelling because ontic vagueness leads to various difficulties, in particular, it either can be dissolved into representational vagueness, or leads to incoherence, or practically unapproachable.

References:

1. Barnes, E. (2008). Indeterminacy, Identity and Counterparts: Evans reconsidered. *Synthese*, 168(1), 81-96. doi:10.1007/s11229-008-9314-x
2. Barnes, E. (2010). Ontic vagueness: A guide for the perplexed. *Noûs*, 44(4), 601-627. doi:10.1111/j.1468-0068.2010.00762.x
3. Evans, G. (1978). Can there be vague objects? *Analysis*, 38(4), 208-208. doi:10.1093/analys/38.4.208
4. Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30(3-4), 265-300. doi:10.1007/bf00485047
5. Lewis, D. (1988). Vague identity: Evans misunderstood. *Analysis*, 48(3), 128-130. doi:10.1093/analys/48.3.128
6. Lewis, D. (1999). Many, but almost one. *Papers in Metaphysics and Epistemology*, 164-182. doi:10.1017/cbo9780511625343.010
7. Sainsbury, R. M. (1995). Why the world cannot be vague. *The Southern Journal of Philosophy*, 33(S1), 63-81. doi:10.1111/j.2041-6962.1995.tb00763.x
8. Williams, J. (2007). Multiple actualities and ontically vague identity. *The Philosophical Quarterly*, 0(0). doi:10.1111/j.1467-9213.2007.538.x

Duties of Friendship: A Refutation of Shiffrin's Transcendental Argument For the Power to Promise

By David Sutherland

University of California, Los Angeles

Duties of Friendship: A Refutation of Shiffrin's Transcendental Argument For the Power to Promise

Abstract

In her articles “Promising, Intimate Relationships, and Conventionalism” and “Immoral, Conflicting, and Redundant Promises,” Professor Seana Shiffrin utilizes a transcendental argument in order to elucidate her rights transfer theory of promising. Transcendental arguments take a proposition that we know to be true and then seeks a necessary condition for that proposition. If there is a necessary condition for that proposition, given that we already know that the proposition is true, it must be the case that the necessary condition is true as well. Transcendental arguments allow us to indirectly prove the veracity of various necessary conditions for known truths, so long as we can prove their conditional nature. Shiffrin uses a transcendental argument in order to argue that we must have the power to promise because the power to promise is necessary to form minimally morally decent relationships and is therefore also necessary to the development of our autonomy.

While Shiffrin’s account of the power to promise is generally compelling, I will argue that her account misses significant complexities inherent in our ability to voluntarily morally obligate ourselves and, thus, her transcendental argument for the power to promise has a far less steady foundation than she acknowledges. I will argue this by first explaining the general formula for transcendental arguments, then introducing Shiffrin’s transcendental argument for the power to promise, then explaining why Shiffrin thinks promises are morally binding, before introducing a duty of loyalty as a tool to reveal some of the complexities in our ability to morally bind ourselves that Shiffrin does not account for. If there are scenarios in which we do not need the power to promise in order to morally bind ourselves, and thus we do not need the power to promise to form

minimally morally decent relationships, then Shiffrin's conclusion that we must have the power to promise seems to lose its footing. Ultimately, I still find Shiffrin's theory on the power to promise to be attractive, but these nuances need to be explained if the theory is to hold its ground against rival theories.

Duties of Friendship: A Refutation of Shiffirin's Transcendental Argument For the Power to Promise

In her articles “Promising, Intimate Relationships, and Conventionalism” and “Immoral, Conflicting, and Redundant Promises,” Professor Seana Shiffirin utilizes a transcendental argument in order to elucidate her rights transfer theory of promising. While Shiffirin’s account of the power to promise is generally compelling, I will argue that her account misses significant complexities inherent in our ability to voluntarily morally obligate ourselves and, thus, her transcendental argument for the power to promise has a far less steady foundation than she acknowledges. I will argue this by first explaining the general formula for transcendental arguments, then introducing Shiffirin’s transcendental argument for the power to promise, then explaining why Shiffirin thinks promises are morally binding, before introducing a duty of loyalty as a tool to reveal some of the complexities in our ability to morally bind ourselves that Shiffirin does not account for.

Shiffirin builds her argument regarding the power to promise on a transcendental argument for that power. In order to fully understand Shiffirin’s argument, we must first understand how a transcendental argument functions. Transcendental arguments take a proposition that we know to be true and then seeks a necessary condition for that proposition. If there is a necessary condition for that proposition, given that we already know that the proposition is true, it must be the case that the necessary condition is true as well. Transcendental arguments allow us to indirectly prove the veracity of various necessary conditions for known truths, so long as we can prove their conditional nature.

Shiffirin’s transcendental argument for the power to promise runs along these same lines. Shiffirin begins with an interesting view of autonomy. She states that due to our nature, our

“embodiment and development” of our autonomy necessarily relies on our ability to form “morally respectful” and “empower[ing]” relationships with others (Shiffrin, “Promising, Intimate Relationships” 520). Shiffrin later refers to these relationships as “minimally morally decent ... relationships” (Shiffrin, “Promising, Intimate Relationships” 499). Shiffrin then asserts that our ability to form minimally morally decent relationships with others relies on our power to promise (Shiffrin, “Promising, Intimate Relationships” 499). If these two premises are taken to be true then, given our ability to fully exercise our autonomy, Shiffrin’s transcendental argument for the power to promise would hold significant weight.

Before we accept her transcendental argument as true, we must first understand why Shiffrin believes that our ability to form minimally morally decent relationships hinges on our power to promise. Shiffrin argues that our power to promise solves vulnerabilities in relationships that arise simply as a result of humans acting as separate and distinct rational actors (Shiffrin, “Promising, Intimate Relationships” 521). These vulnerabilities not only damage the moral nature of the relationship, but they can also damage the relationship itself. Solving these vulnerabilities is necessary for us to form minimally morally decent relationships.

Shiffrin offers a compelling example of two friends moving to a city together (Shiffrin, “Promising, Intimate Relationships” 504) to elucidate why these vulnerabilities can exist, why these vulnerabilities make minimally morally decent relationships impossible, and how our power to promise can solve these vulnerabilities. I will add small details to this thought experiment in order to elucidate the points Shiffrin hopes to make. Say that two friends, Friend A and Friend B, want to move to Portland together. Friend A’s dream job is in Portland. She would also like to live together with Friend B in Portland. However, her financial state is such that she would be incapable of

moving to Portland alone. Friend B would also like to live with Friend A in Portland. Friend B has a job offer for forty-thousand dollars per year in Portland. However, Friend B also has a job offer for sixty-thousand dollars per year in New York City. Friend B is financially capable of moving to New York City by herself. Despite this, both friends' ideal scenario would be to move to Portland with the other. In this example, Friend A is more vulnerable than Friend B due to the fact that she cannot move without B and because B has a better job offer somewhere else. If we analyze what would develop in this example without the power to promise, we will learn why Shiffrin believes that this vulnerability must be solved in order for Friend A and Friend B to maintain a minimally morally decent relationship.

First, this vulnerability may damage the moral standing of the relationship. In a world without the power to promise, Shiffrin stipulates that it would be impossible for either party in this scenario to morally bind themselves to moving through a declaration of intention (Shiffrin, "Promising, Intimate Relationships" 504). In this scenario, A may feel compelled to "sweeten the pot" (Shiffrin, "Promising, Intimate Relationships" 505) for B in order to further entice B to move to Portland. Sweetening the pot may include the transfer of goods and services from A to B on condition that B moves to Portland with her. Shiffrin believes that sweetening the pot would be a "concrete realization of their disproportionate vulnerability," and would "replace the inequity of vulnerability with a more substantive inequity — a transfer of goods or services" (Shiffrin, "Promising, Intimate Relationships" 505). Sweetening the pot, Shiffrin argues, is an instance of exploitation and is therefore morally reprehensible.

Furthermore, the disproportionate vulnerability between the friends may destroy the relationship itself. Her vulnerability may lead A to feelings of "powerlessness or frustration that can

further make the relation more fraught” (Shiffrin, “Promising, Intimate Relationships” 504). The vulnerability can also lead A to seek constant assurance that it is still B’s intention to move to Portland with her, which could also fray the relationship. If the vulnerability is prolonged, the relationship itself may suffer as a result.

So long as B cannot morally bind herself with respect to moving to Portland these inequalities and vulnerabilities will remain. This is because A would be aware that B would be morally free to change her mind at any moment. However, Shiffrin argues, if B could promise to A that she will move to Portland with her, B would no longer be morally free to not move to Portland, which would forestall the troubling developments that occur due to the vulnerability. There would be no feelings of powerlessness, no need to sweeten the pot, and no need to seek constant reassurance. The power to promise acts as a “conversation-stopper” (Shiffrin, “Promising, Intimate Relationships” 506) in this way. This conversation stopping aspect of the power to promise is why the power to promise solves the vulnerabilities that would otherwise make minimally morally decent relationships impossible. Thus, the power to promise is necessary to form minimally morally decent relationships.

One may wonder why B is no longer morally free to change her mind after making a promise. Shiffrin offers a “rights-transfer view” of promising in order to explain this (Shiffrin, “Immoral, Conflicting” 155). According to this theory, a promise involves a valid transfer, from the promisor to the promisee, of the right to decide whether and how a specified action will be taken (Shiffrin, “Immoral, Conflicting” 156). Once a promise has been made, the promisor no longer has the moral right to decide to act otherwise because they transferred that right to the promisee

(Shiffrin, "Immoral, Conflicting" 157). This explains why it is immoral to break promises and, in turn, why they are morally binding.

While I find this view of promising to be somewhat convincing, I worry that Shiffrin's transcendental argument for the power to promise loses weight due to the fact that she overlooks the complexities involved in how we can voluntarily morally obligate ourselves. My instinct is to deny her stipulation that *all* voluntary moral obligations are consistent with the power to promise.

One way we can discover the nuances that Shiffrin glosses over in her account of the power to promise is to entertain the idea that there may be a duty of loyalty between friends that can also morally bind them. Under this view, people would be able to choose who they are friends with and what kind of relationship that friendship is. They would then incur moral obligations in relation to how their relationship functions with respect to the other person.

I would argue that a duty of loyalty could function similarly to the power to promise. If we return to the original case Shiffrin proposed, in which both Friend A and Friend B want to move to Portland but Friend B cannot use the power to promise to morally bind herself to moving, I would argue that invoking a duty of loyalty would also have a conversation-stopping effect. If Friend B said, "I am your friend, so of course I will go to Portland with you. That is what we do for each other," Friend A would no longer have the need to sweeten the pot or seek further reassurance. If Friend B asserted that she was good enough friends with Friend A to move to Portland with her, and thus incurred the obligations inherent in their relationship, then the vulnerabilities Friend A faces would be solved. To me, it seems that the duty of loyalty could achieve similar effects as the power to promise, and thus Shiffrin paints with too broad a brush when she stipulates that without a power to promise Friend B loses all ability to voluntarily morally bind herself.

Shiffrin's main objection to this account may be that the statements Friend B could make in an attempt to invoke a duty of loyalty are actually just implicit forms of promises. Shiffrin asserts that the power to promise is constituent of a spectrum of promises — including implicit, informal, and formal promises (Shiffrin, "Promising, Intimate Relationships" 514). When Friend B says, "I am your friend, of course I will go to Portland with you," Shiffrin would argue that she is actually making an implicit promise to go to Portland with Friend A. I am inclined to deny that this is the case. Let us go back to the Portland example to see why. If Friend B promised to Friend A that she will move to Portland, and then proceeded to renege on her promise, Friend A's reaction would be, "Friend B broke her promise." However, if Friend B instead stated, "I am your friend, of course I will go to Portland with you," and then proceeded to go back on her word, Friend A's reaction would not be "Friend B broke her promise." Instead, it would be more along the lines of, "I guess Friend B is not as good of a friend as she said she was." This seems to me to suggest that Friend B, in invoking their friendship, was not making an implicit promise, but rather successfully invoking their friendship as a morally binding force. If nothing else, this peculiar reaction belies one important complexity in Shiffrin's account of the power to promise that she does not touch on in her articles.

This complexity regarding our ability to voluntarily morally obligate ourselves is significant because it decreases the weight behind Shiffrin's transcendental argument for the power to promise. Shiffrin's transcendental argument rests on the premise that the only way we can voluntarily morally bind ourselves to a statement of intention is through the power to promise. This is how she is able to argue that the power to promise is necessary to form minimally morally decent relationships, and minimally morally decent relationships are necessary to fulfill our autonomy, and thus, given that we are capable of fulfilling our autonomy, we must have the power to promise. If

there are scenarios in which we do not need the power to promise in order to morally bind ourselves, and thus we do not need the power to promise to form minimally morally decent relationships, then Shiffrin's conclusion that we must have the power to promise seems to lose its footing. Ultimately, I still find Shiffrin's theory on the power to promise to be attractive, but these nuances need to be explained if the theory is to hold its ground against rival theories.

Bibliography

Shiffrin, Seana. "Promising, Intimate Relationships, and Conventionalism." *SSRN*, 28 Apr. 2016, papers.ssrn.com/sol3/papers.cfm?abstract_id=2771375.

Shiffrin, Seana. "Immoral, Conflicting and Redundant Promises." *Reasons and Recognition Essays on the Philosophy of T.M. Scanlon*, www.academia.edu/3428904/Immoral_Conflicting_and_Redundant_Promises.

Belief Polarization: Is Path-Dependent Evidence Search Irrational?

By Shuyuan Wang

University of Chicago

Belief Polarization: Is Path-Dependent Evidence Search Irrational?

I. Introduction

In his “Disagreement, Dogmatism, and Belief Polarization,” Thomas Kelly presents his descriptive model of the psychological tendencies underwriting belief polarization, one that serves as an alternative to the Kripkean Dogmatism model. Kelly points out that, instead of immediately *dismissing* evidence that counts against one’s prior belief, one tends to 1) *increase scrutiny* of such piece of evidence, and 2) devote more cognitive resource *search for alternative explanations* for it. Kelly argues that, whereas Kripkean Dogmatism violates the Commutativity of Evidence Principle, the psychological tendencies described in Kelly’s Alternative Model does not; then, although the latter psychological tendency is still characterized by an inevitable element of *path-dependence* in evidence acquisition—one that is far from ideal—the latter tendency is still rational. In what follows, I will argue that the psychological tendency in Kelly’s model is more rational than Kripkean Dogmatism, despite the apparently unsettling implications of such path-dependence.

I will begin by providing an overview of the two models, Kripkean Dogmatism and the Alternative Model, as well as Kelly’s assessment of the normative implications of both models. Then, I will consider a tempting line of reasoning that problematizes such path-dependence: the tendency posited by Kelly’s model brings about a body of subsequently-attained evidence *skewed* in a direction in favor of one’s prior belief, and this fact amounts to an instance of *biased search* sufficient to count the subject as *irrational*—or even no less irrational than a Kripkean Dogmatist. Finally, I will offer a reply to the above argument: the normative standard for placing such additional requirement of rationality is, in fact, extremely unclear; under a proper setting of the normative standard for rationality, the agent exercising the psychological mechanism is still rational.

II. The Two Models, Explained

To begin with, both models serve to explain the phenomenon of *Belief Polarization*: when two subjects holding opposing beliefs encounter a common body of mixed evidence, the disagreement becomes even more pronounced, rather than being mitigated.¹ Two sorts of questions are worth asking, one *descriptive* and one *normative*:

1. *Descriptive*: What is the psychological mechanism underwriting such phenomenon?
2. *Normative*: Is it possible that such psychological mechanism is not just blatantly irrational?

The Kripkean Dogmatism Model is a preliminary with which Kelly opens the discussion of the descriptive and normative considerations. Particularly, Kripkean Dogmatism describes a sort of reasoning in which one uses one's prior belief as a license to dismiss counterevidence as *misleading evidence*.² Here is an example to illustrate how it contributes to belief polarization:

Suppose you and I are both Kripkean Dogmatists. I believe in the factual claim that *death penalty deters murder*, while you believe that such proposition is false. From my belief that the proposition is true, it follows that that *any evidence that suggests that death penalty has no effect deterring murder is misleading*. Since I am concerned to access the truth of the issue, dismissing misleading evidence would be a sensible policy for me to adopt. Hence, when encountering a mixed body of evidence, I dismiss the portion of it that counts against my belief as misleading. Meanwhile, there is no reason for me to dismiss the portion that supports my former belief, and I might very well let it confirm my former belief. Also, you reason in a parallel and opposite way—as a result, our beliefs are further apart.

According to Kelly, it is uncontroversial that Kripkean Dogmatism is irrational. One way to see such irrationality unfold is that Kripkean Dogmatism violates the Commutativity of Evidence Principle. The principle says that, to the extent that what one is justified in believing in depends on one's body of total evidence, the temporal order in which particular pieces of

¹ It is natural to expect that evidence should somehow serve as an intersubjective neutral arbiter (See Kelly, 2016)—that is, evidence is frequently linked to the desideratum of *objectivity*, and it is expected to mitigate the disagreement arising out of the limited subjective perspectives of the subjects. This could be a common and attractive thought underlying how rationality is expected to be in line with neutrality, or a convergence after such common exposure to evidence.

² A piece of evidence *E* is *misleading* if *E* suggests proposition *P* and *P* is false.

evidence is attained should make *no* difference to what one is justified in believing in³. To picture how such violation takes place, consider a scenario in which the total body of evidence consists of two pieces, E1 suggesting P, and E2 suggesting \sim P. Each piece is credible and strong enough to let the subject form a strong belief in the suggested proposition. Then, for an agent who started out neutral, the order in which the subject obtains E1 and E2 leads to different resultant belief: having received E1, the subject forms a strong belief in P, which enables her to dismiss E2 that came subsequently and still hold P; if E2 came earlier than E1, she would end up with a belief in \sim P. The Commutativity of Evidence Principle is violated, blatantly.

Kelly pointed out that, luckily, it is unlikely that we are Kripkean Dogmatists. Empirical studies show that we tend to devote significantly more attention to counterevidence, rather than just dismissing it. In this light, Kelly introduces his Alternative Model, highlighting two closely connected psychological tendencies. Firstly, one tends to *increase scrutiny* of counterevidence and become more apt at identifying its flaws. Secondly, given such additional attention, one is disposed to devote more cognitive resource to *search for alternative explanation* that purports to explain the counterevidence. To see how the psychological tendency contributes to polarization, consider the same example concerning *you and I disagreeing on deterrence*:

Suppose we both reason according to Kelly's model, and we are both exposed to a common body of mixed evidence. You have the prior belief that *death penalty does not deter murder*, while I believe in the opposite. Upon encountering statistical report that suggests the contrary hypothesis—say, that *state A has death penalty while state B doesn't*, and that *state A has significantly lower murder rate*—you would devote more attention to this report. You scrutinize it more, and you are cognitively more apt at identifying its methodological flaws, in virtue of your prior belief. If it *is* the case that you have found a genuine flaw⁴, you can then reasonably dismiss this piece of counterevidence. Also, you are likely to consider alternative explanations for this report, and to the extent that you have actually found formidable alternative explanations⁵, you would not increase your credence in the *deterrence* hypothesis as much as I

³ Kelly, *Disagreement, Dogmatism, and Belief Polarization*, p7. I think it would be a good question to ask whether the Commutativity of Evidence Principle should be accepted as uncontroversial; however, within the scope of discussion of this paper, I will assume that the Commutativity of Evidence Principle is true.

⁴ Note that this increased scrutiny is not driven by one's intention to defeat the opposing position—that would surely be irrational—but a natural devotion of some extra consideration onto counterevidence. It is arguably natural to devote more attention to evidence that does not cohere with one's currently-held belief system, for instance, just for the sake of learning more.

⁵ Similarly, the search for alternative explanation is not done deliberately for some *ad hoc* hypothesis, and it can be natural in one's reasoning process.

would, as you are aware of alternative explanations that I am unaware of. Since what one is justified in believing in also depends on the space of potential alternative explanations, the support that the counterevidence lends to the hypothesis it suggests will get diluted by the presence of plausible alternative hypotheses.

According to Kelly, the tendency described in the Alternative Model does not violate the commutativity of evidence principle. Indeed, under this tendency, the temporal order in which evidence is acquired *does* alter the belief one arrives at; but in the meantime, the total body of evidence one ends up with will also not remain the same. There can be two senses of ‘evidence’: evidence in the *narrow* sense just consists of relevant factual information about the world, for instance, statistical data relevant to the study of whether *death penalty is a deterrent*. Evidence in the *broad* sense consists of anything that would make a difference to what one is justified in believing; in our case, *broad evidence* include both the narrow evidence and the *space of relevant alternatives* one is aware of. Then, the two subjects end up with different bodies of evidence in the broad sense, as there exists a difference in the possible alternatives they are aware of. In this sense, the temporal order in which evidence is acquired can make a causal difference to which body of total evidence one ends up with, and our belief formation is characterized by path-dependence. But it is still true that the beliefs of both subjects are accurately proportioned to the bodies of evidence they have, and the Commutativity of Evidence Principle is respected: for each subject, given the total body of evidence one has remains the same, and that the order in which constituent pieces are received changes, what one would be justified in believing in would still remain the same—the belief that is accurately proportioned to the body of broad evidence that *this* subject is aware of.

In my opinion, the fact that the psychological tendency brings about a different body of *broad* evidence is worth focusing on. Kelly recognizes that, the tendency is going to render the total body of evidence one ends up with *biased in a systematic way*, that is, skewed in a direction in favor of one’s former beliefs. I take that the *search for alternative explanation* as the major contributor to such biasing mechanism: one receives more *broad* evidence, that is, formidable alternative explanations, along the way of their inquiry; in a sense, the alternative explanations are still in favor of one’s prior beliefs, so long as they dilute the support received by the

opposing hypothesis. Kelly admits that, so long as one is aware of such mechanism, one ought to correct for it by lowering the credence in the belief one eventually arrives at. Furthermore, the acquisition of broad evidence depends on the historical order in which pieces of early evidence is received—in this sense, what body of *broad* evidence one could end up with becomes a contingent, arbitrary matter. Then, we will need to question whether one's exercising of such biasing mechanism ought to be categorized as irrationality.

III. A Tempting Thought: Irrationality Lies within the Skewing Mechanism

Here is a tempting reason to think that the psychological tendency in the Alternative Model is irrational: the fact that one exercises such biasing mechanism that brings about a skewed body of evidence speaks sufficiently of the subject's irrationality. Such biasing mechanism enables the historical fact about the temporal order in which evidence is received to implausibly affect *the subsequent search and exposure to further broad evidence*. In this sense, we could argue that the subject violates a somewhat extended version of the Commutativity of Evidence Principle and, hence, we can conclude that the subject is nearly as irrational as a Kripkean Dogmatist.

To begin with, to the extent that the *result* of such systematic biasing mechanism—the skewed body of evidence—is concerning, our normative standard of rationality should capture such issue and problematize the exercising of the mechanism. This means that the normative standard Kelly put forth needs to be heightened, as it is limited in a significant respect—it does not problematize the *biased search*.

In Kelly's account, accurately proportioning one's belief to one's total body of evidence suffices for believing rationally.⁶ If one is *aware* of the presence of the biasing mechanism and, hence, knows the direction in which one's body of evidence is skewed, then one ought to correct for it by lowering one's credence—not doing such correction would surely be irrational.⁷ But this does not mean that any ordinary person who is *unaware* of it is *irrational* in exercising

⁶ Ibid. p23.

⁷ That is, the factual information—in other words, evidence—about such biasing mechanism occurs within one's scope of awareness, and ignoring this piece of information would mean one's failure in proportioning one's belief to the body of evidence.

such mechanism. After all, his belief is still accurately proportioned to his own body of *broad* evidence; also, to reiterate the normative implication, the Commutativity of Evidence Principle is still respected in this person's case—any change in the order in which the constituent pieces of *this person's* entire body of broad evidence is received will not make a difference in what he is justified in believing in.

One could well disagree with Kelly's normative standard here. Here, the kind of rationality that Kelly is concerned with is merely the mapping relation between a) the body of evidence one is aware of, and b) one's belief. By this standard, so long as one's belief is *accurately proportioned* to the evidence, one is rational (or not irrational, at least). But clearly, it does not cover *how* one's body of evidence is arrive at, that is, the *legitimacy of search*, an issue we identified clearly in the biasing mechanism. After all, Kelly's Alternative Model speaks of an important fact: one's responding to counterevidence is not just a simple matter of passively forming beliefs as a 'function output' of evidence; a large part of the process involves the *activity* of searching. It seems implausible that a sensible requirement of rationality should omit giving constraints on such a big, important part of one's reasoning process. So long as one's own cognitive agency participates in creating such biasing mechanism, one is culpable for exercising a biased search. Hence, we need to heighten the standard for rationality: responding to counterevidence in a way that brings about a skewed body of evidence is irrational. This standard holds true, regardless of whether the subject is *aware* of the fact of such mechanism, or not. A case in which one puts such tendency to exercise, while being *unaware of* it, is just a case of unreflective biased search. The ideal case of respecting such normative standard is someone who does not create the skew, or the imbalance, in the acquisition of subsequent broad evidence. Someone who exercises such biased search but lowers her credence afterwards would fall into a less-than-ideal case: the irrational mechanism is still exercised, but the agent subsequently corrected for it.

It follows that Kelly's application of the Commutativity of Evidence Principle does not go beyond the mapping relation between evidence and belief, and it is not able to help us identify biased search. But as the normative standard of rationality is revised, we can problematize the time-dependent nature of evidence acquisition in Kelly's model, particularly caused by the

search for alternative explanation. As Kelly pointed out, the temporal order in which evidence is acquired can make a causal difference to which body of total evidence one ends up with. the acquisition of broad evidence has thus become a contingent, arbitrary matter, as one could totally end up with a different body of evidence. In a sense, the original commutativity of evidence principle problematizes the time-dependent phenomenon of one's belief formation: rationality is in question *when one's belief ends up differently as a result of the difference in the order of evidence acquisition*, when the total body of evidence is the same. This speaks of the problem of a Kripkean Dogmatist. To extend the principle a bit more, rationality is in question *when one's belief ends up differently, as a result of the difference in the order of broad evidence acquisition, which is an arbitrary result of the difference in the order of narrow evidence acquisition*.

In short, the above tempting though suggests that, to the extent that the biasing mechanism is concerning, our normative standard of rationality should be set in a way that problematizes it. Under the extended normative standard for rationality, the subject reasoning under the Alternative Model is irrational in a way somewhat analogous to a Kripkean Dogmatist.

IV. A Reply: Searching for Alternative Explanation as a Sensible Cognitive Policy

The thought above is tempting. However, through closer examination, we will see that raising the normative standard according to the suggestion above involve tremendous difficulties and ambiguities—in fact, problematizing the biased search process has problematic normative implications on its own. We will focus on the search for alternative explanation here, the primary contributor to the disproportionate amount of broad evidence in one's favor. Searching for alternative hypothesis—or putting some minimal preference in one's prior beliefs—may be a sensible, or even indispensable policy to adopt in face of potentially misleading counterevidence, the abandoning of which will cause problem on its own. That is, we cannot simply claim that the tendency described by Kelly's model to be blatantly irrational.

Suppose that we do follow the suggestion given in the last section and heighten the standard for rationality. Imagine a rational agent who meets up to the standard perfectly, that is, she consciously adopts a policy that counteracts any tendency that would result in a

disproportionate increase in broad evidence in favor of her side—she will not put the hypothesis of the opposing side in any disadvantage. Say, she may not exercise the search for alternative explanations at all, or actively suppress such tendency. If she does come up with ideas about alternative explanations for a counterevidence, she would also devote just as much cognitive resource to activities that ‘balances the game,’ say, 1) to scrutinize the existing evidence that supports her prior belief or 2) try to come up with alternative explanations for those evidence.⁸ In general, she regulates herself in such a way that the subsequent body of broad evidence she encounters does not have any sort of imbalance. In this sense, she highly appreciates the value counterevidence, and she does not let her prior belief take any part in undermining the support that the counterevidence gives to the opposing hypothesis. She may well let the newly-encountered counterevidence defeat some parts of her earlier-held belief, or lower her credence. This agent may be conceived of as an ideal type of a flexible, neutral, and open-minded thinker.

But the policy that the agent adopts has its own problem: it does not guard her against misleading evidence. In a sense, increasing one’s scrutiny or searching for alternative explanation for a piece of evidence is a measure to account for the possibility that evidence might be misleading, and—if it turns out to be misleading—to mitigate the effect of the misleading evidence, that is, not to be seriously misled by misleading evidence. Kripkean Dogmatists’ dismissing of counterevidence has a similar rationale, although its practice is somewhat radical. Then, exercising a psychological tendency with a reasonable preference of one’s prior beliefs means that one recognizes that the encounter with misleading evidence is not entirely impossible. Indeed, misleading evidence is less likely to occur than evidence that suggests truth, but still, one has to have a back-up plan to deal with misleading evidence in case it does occur. On the other hand, an ideally flexible, open-minded thinker has no way to guard against misleading evidence. If it turns out that the piece of counterevidence she encountered is misleading, she would not undermine it in any way—she does not ‘explain it away’—and

⁸ The policy that this rational agent adopted may not be the only way to balance out the effect of the biasing mechanism...but at least, the above postulation serves to give a demonstration of how a minimal amount of ‘self-biasing’—or some preference given to the side of one’s prior beliefs—is necessary for someone who is concerned about accessing truth in face of misleading evidence.

she may well let it lower her current credence, more than it should be. So long as one is concerned with accessing the truth of the issue, while acknowledging the possibility of encountering misleading evidence, one would not insist on doing nothing to mitigate the effect of counterevidence.

Indeed, tendencies such as searching for alternative explanations would lead to an unsettling result: it brings about a skewed body of *broad* evidence. But wholly categorizing it as irrational would lead to the implausible consequence described above. In general, there is a minimal degree of self-preference that should be given, in light of the possibility that evidence might be misleading. This is not to say that any kind of self-preference should be justified—that would surely be irrational—but some occasional implementation of self-preference policy is rational, or even necessary. The bar for rationality should not be heightened, and Kelly's formulation suffices for the purpose of the discussion of rationality in an ordinary sense, not in the ideal sense. The tendency to search for alternative explanation is highly interlinked with such self-preference policy. It would be hard to set in stone the precise line past which such self-preferencing search for alternative explanation should be irrational, but in the neutral cases presented by Kelly, where such search is natural and not ad hoc, the agent can be called rational. Of course, the biasing mechanism is far from ideal, but some neutral cases of it suffices for the standard of ordinary rationality. An agent who does not *consciously* adopt this policy—that is, being unaware of the mechanism—but keeps in mind the possibility that evidence might be misleading, may well deserve to be called rational (at least, not be categorized as irrational).

V. Conclusion

In this essay, I have argued that the psychological tendencies described in Kelly's Alternative Model is more rational than Kripkean Dogmatism, although it is far from ideal. I have anticipated a line of argument that the psychological tendency in Kelly's model is irrational, and that the standard for rationality should be raised to capture the *biased search*. I have offered a reply to such line of reasoning: because of the normative ambiguity within adopting a biased search policy, simply categorizing it as irrational will be problematic; hence,

under a more proper setting of normative standard, the subject in Kelly's model deserves to be called rational.

nyuaupc.org

The Annual Undergraduate Philosophy Conference is organized by the NYU Philosophy Forum.

Visit us at

- nyuphilforum.org
- as.nyu.edu/philosophy/undergraduate/philosophyforum

All articles published in the AUPC Journal are fully open access: immediately freely available to read, download and share. All articles are published under the terms of the Creative Commons Attribution License. The Creative Commons Attribution License permits use, distribution and reproduction in any medium, provided the original work is properly cited and allows the commercial use of published articles.