LOGOS

The Cornell Undergraduate Journal of Philosophy

Volume XI - Spring 2014

Logos : λόγος

Volume XI • Spring 2014

An independent student publication

Logos: The Undergraduate Journal of Philosophy, an independent student organization located at Cornell University, produced and is responsible for the content of this publication. This publication was not reviewed or approved by, nor does it necessarily express or reflect the policies or opinions of, Cornell University or its designated representatives.

Mail:
Logos: The Undergraduate Journal of Philosophy
c/o The Sage School of Philosophy
218 Goldwin Smith Hall
Cornell University
Ithaca, NY 14853

Email:
journal.logos@gmail.com

Logos : λόγος

Volume XI • Spring 2014

Contents

Editors' Introduction

The staff of *Logos* is proud to present the eleventh volume of Cornell University's undergraduate journal of philosophy. After carefully considering the submissions we received over the past year we have selected an exemplary set of five articles chosen for their creativity, cogency, and depth of philosophical inquiry.

This year's selection pool was full of quality submissions, and we received inquiries from over eighty undergraduates situated across the English-speaking world. All of the papers contained within this volume were carefully reviewed and selected because of their exceptional quality and varied subjects. The eleventh volume of *Logos* features papers whose topics fall under the headings of philosophy of physics, ethics, ancient philosophy, philosophy of religion, and free will. We are delighted to be able to publish such a broad set of articles while bringing the best new undergraduate work to public view.

We would like to thank and acknowledge the authors of our chosen submissions: Alexander Meehan for his submission entitled "Quantum Mechanics without Math: Why We Must Nominalize the Dynamical Role of Chance," Dallas M. Ducar for his submission entitled "On the Reason to Reduce the Efficacy of Moral Luck," Michael Allen Ziegler for his submission entitled "Wily Socrates: How Seriously Should We Take the Arguments of Hippias Minor?," Matthew Duvalier McCauley for his submission entitled "The Problem of Theological Fatalism," and Matthew Paskell for his submission entitled "Manipulation, Argument, & Experiment: Putting Folk Intuitions into Context."

We are indebted to Professor Ted Sider for leading a thoughtful talk on metaphysics at our discussion club this past fall; to the staff of the Sage School of Philosophy for assisting with publication, the Life Raft Debate, and the day-to-day of running the journal; and to our undergraduate staff without whom none of this would be possible. We are grateful to the Student Assembly Finance Commission whose funding supports *Logos*, and to our advisor Professor William Starr, for his continued support and willingness to go above and beyond in making our goals reality.

Daniel Cook
Editor-in-Chief

Sadev Parikh
Club President

Quantum Mechanics without Math:

Why We Must Nominalize the Dynamical Role of Chance

Alexander Meehan
Brown University

1. INTRODUCTION

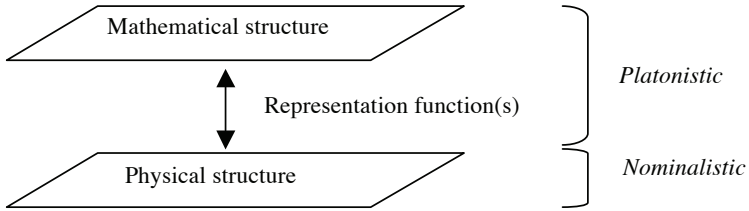
Is mathematics indispensable to the formulation of our fundamental physical theories? In his 1980 monograph *Science Without Numbers*, Hartry Field claims the answer is *no*: we can 'nominalize' our physical theories, i.e. reformulate them so that they do not assume the existence of any abstract mathematical objects (e.g. numbers), or the truth of any mathematical theorems (e.g. the fundamental theorem of calculus). Field argues that math can be thought of as a 'conservative extension' over our physical theories; it does not allow us to reach any conclusions about physical reality beyond what we can, in principle, reach without its help (Field 1980, pg. 11). Its logical consistency, however, *does* make it a useful and reliable tool for moving efficiently from premises to conclusions in our physical reasoning.

Field's motivation for carrying out this nominalization process is to combat a strong counter-argument to Anti-Platonism. Platonism is the view that there exist abstract mathematical objects, and that these objects are aspatiotemporal and acausal; they do not interact at all with the physical world (call this the Principle of Causal Isolation, or PCI). Under Platonism, mathematical theories are claims about such objects and are taken to be literally true. According to *Anti*-Platonism, we need a different account of math because there are simply *no such things* as mathematical objects. Field endorses a particular kind of Anti-Platonism which holds that mathematical theorems and objects are fictional—so sentences like "2+3=5" have the same status as "Sherlock Holmes lives in 221B Baker Street."[1] This kind of Anti-Platonism is called Fictionalism. The popular counter-argument to Anti-Platonism that Field wants to combat is known as the 'Quine-Putnam indispensability argument'. This argument takes as its premise that the application of mathematics is indispensable to our fundamental physical theories. It points out that we take these physical theories to be true and well-justified; thus, if we are to buy into those theories, we should also buy into the truth of the mathematical theorems and the existence of the mathematical objects that are indispensable to them.

Field's nominalist project is an attempt to demonstrate, or at least indicate in a convincing way, that the premise outlined above is false. His general strategy is to show that he can define a physical structure using metaphysically primitive relations formalized into predicates of a logical language, and that he can then mathematically map this (nominalistic) physical structure onto the physical theory's (platonistic) mathematical structure. These maps are 'representation

[1] Therefore, "there exists a prime number between 3 and 5" is trivially false because numbers do not exist, and "there does not exist a prime number between 3 and 5" is trivially true for the same reason.

functions' and they typically take the form of homomorphisms.[2] Once these representation functions are defined, we can use them to eventually prove that the mathematical structure does not entail any consequences that cannot be stated nominalistically at the level of the physical structure and is therefore, in principle, dispensable.[3] Note that for the nominalist (and, by extension, the fictionalist), the representation functions are fictional, just like the mathematical structures they map to; she simply develops them to convince the Platonist of the mathematical structure's dispensability.

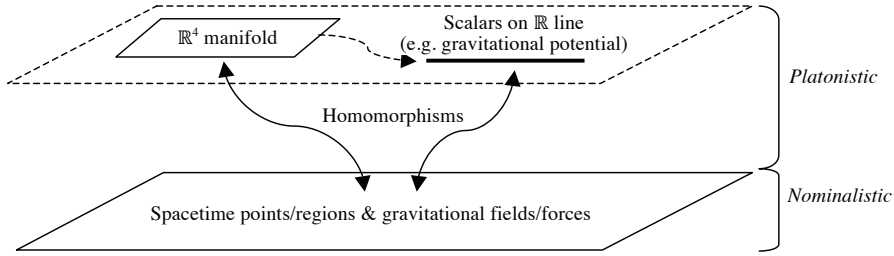


Using the strategy outlined above, Field effectively manages to nominalize Newtonian Gravitation Theory (NGT). NGT is a classical field theory that represents the gravitational force by a mathematical field satisfying certain differential equations. The field is defined on a four-dimensional manifold, the points of which represent physical points of spacetime. In Field's nominalization, the primitive relations are formalized into a language whose quantifiers range over spacetime points/regions (he somewhat controversially takes spacetime points/regions to physically exist in a literal sense—this view is known as substantivalism). The basic idea is that Field makes use of these primitive relations (e.g. 'betweenness' and 'congruence') to talk of distances between spacetime points, and scalar values (e.g. gravitational potential) that belong to spacetime points, without actually using numbers. With the help of a rich logical apparatus, he continues to define relations and prove homomorphisms between the physically real spacetime structure and the mathematical manifold, until eventually he has the power to make statements using his axiom systems that, if true, entail that the platonistic formulation of NGT is also true, and vice-versa. Now the nominalistic and the platonistic formulations have the same nominalistically-statable consequences, and we can correctly consider the mathematical entities dispensable to NGT.

---

[2] A structure-preserving map between two algebraic structures. In simple cases, the representation function is usually an isomorphism, a homomorphism that has an inverse.

[3] I am leaving out some details. For example, in addition to the representation functions (formally stated in "representation theorems") Field argues we may need *uniqueness theorems* which specify the kinds of transformations that, when applied to a representation function, yield another representation function. For an example see (Field 1980, pg. 50-51).



In this paper, I will examine contemporary debates surrounding the nominalization of one of our most well-tested and fundamental physical theories, Quantum Mechanics (QM). I will argue that certain problems one encounters in this nominalization effort, particularly in the nominalization of the role of chance in the 'dynamics' of QM (how quantum systems behave over time) ought to concern *both* the Anti-Platonist and the Platonist. They ought to concern the Anti-Platonist because she will be unable to capture the dynamics of the quantum world by merely listing off the 'nominalistic content' of QM; this is an inability that, as I will explain, has much deeper consequences than the inability to provide a general nominalized theory of QM. They ought to concern the Platonist because, even granted the truth and existence of probability laws and chances as abstract entities, it seems she cannot use them to describe the dynamics of the quantum world without also violating PCI. My central thesis is that if we desire a complete picture of the dynamics of the quantum world that does not violate PCI, then we must seek a metaphysics according to which the role of chance is nominalizable.

2. IS IT POSSIBLE TO NOMINALIZE QM?

In an influential 1982 review of *Science Without Numbers*, David Malament expresses doubts that the nominalist program can extend to QM (Malament 1982, pg. 533-534). He notes that the representation function would probably have to take the form of a homomorphism to a Platonic lattice of subspaces of Hilbert space (in the next paragraph I will explain these technical terms). The only relevant physical structure he can think of that is homomorphic to this Platonic lattice is a 'lattice of *quantum events*'. Yet, Malament's argument goes, unlike spacetime points—which can plausibly be considered physically real[4]—quantum events are surely *abstract* entities that do not literally exist in the physical world. And so, in general, it seems it will be difficult to come up with a theory of what is happening at the physical level that uses *only* non-

[4] Although Field's assumption of substantivalism (that spacetime points physically exist in a literal sense) is not uncontroversial, it is generally taken seriously in contemporary literature.

abstract entities & properties and that can *also* be correctly mapped to the mathematical structure of QM. In order to understand this worry, let me give a rough idea of what this talk of 'quantum events' means. Please note that while the subsequent explanation is relevant, the reader does not need to understand all the mathematical details in order to follow my later arguments.[5]

In the standard formalism of QM, we use vectors in Hilbert space to represent the possible states of a system, and Hermitian operators to represent the observables of a system. In the orthodox interpretation of QM (the interpretation that most contemporary physicists accept, sometimes referred to as the Copenhagen interpretation), we say that 'measurements' of an observable (e.g. position, spin) on a system 'collapse' the general state of the system into one particular state corresponding to the outcome we observe. Which state is collapsed into is a matter of objective probability or *chance*; I say 'objective' because, before the measurement, the system was in superposition—there was *no fact of the matter* about which particular state it was in (e.g. a particle was in a superposition of being in Position X and Position Y or, more familiarly, Schrödinger's cat was in a superposition of being both dead *and* alive—it was not dead, it was not alive, it was not both, and it was not neither). This is distinct from *epistemic* probability, where we may not have epistemic access to the state of the system but there is a fact of the matter.[6] In this paper, I will use 'probability' and 'chance' to mean objective probability.

The set of possible outcomes of the measurement is represented mathematically by a particular set of real numbers. We denote the physical *event* of a measurement of observable A yielding a value in the set of real numbers Δ by (A,Δ). This event can be represented mathematically by $CS(A,\Delta)$, a closed subspace of the Hilbert space H in which A is represented, defined such that a vector v of H is in $CS(A,\Delta)$ *iff* there is a probability of 1 that a measurement of A on a general state represented by v (call it Ψ) will yield a value in Δ. We can then think of Ψ as assigning to each event (A,Δ) a real number r in [0,1] where r is the probability that the event will occur if A is measured on a Ψ-state system. To calculate r we take the inner product of v with the projection of v onto $CS(A,\Delta)$ (Balaguer 1996, "Towards…" pg. 214-215).[7]

Now we define S(E), the set of events (A,Δ) associated with a maximal class of mutually incompatible observables[8], and S(H), a set of closed subspaces in H in which that class of observables is represented. It turns out that S(E)

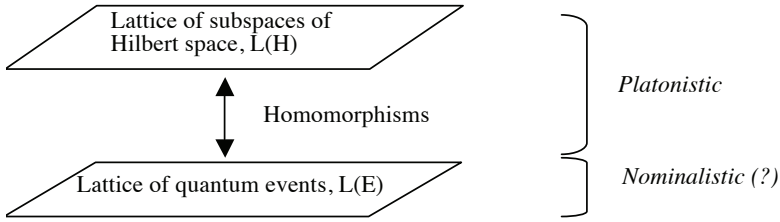
[5] The adventurous reader will find a more technical version of my explanation in (Balaguer 1996, "Towards…" pg. 214-217).

[6] For example, if we know all the initial conditions of a dice roll we can predict the result; normally, however, we do not have epistemic access to this information, so we talk in terms of probabilities instead.

[7] Again, it is not critical that the reader understand these more technical details.

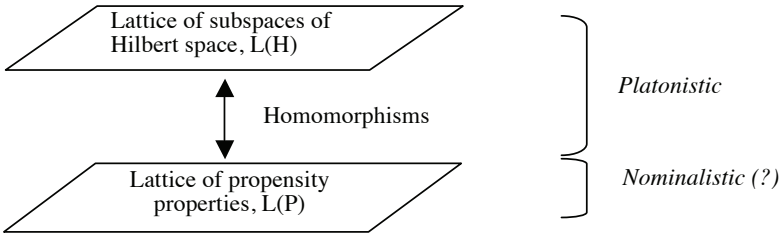
[8] Incompatible observables cannot be simultaneously measured (e.g. position and momentum). 'Maximal' means there are no observables incompatible with all the observables in the set that are not themselves in the set.

and S(H) are not only in 1-1 correspondence, but that we can construct *orthomodular lattices*, sets with a special type of ordering,[9] out of each of them—L(E) and L(H)—which are homomorphic to each other (Balaguer 1996, "Towards…" 216).[10] However, there is a catch: in order to get the full lattice structure of L(E), and thus the homomorphism, we need to make use of *all* of the events of S(E), *many of which have not yet occurred*. Thus, even though we have a homomorphism between L(E), the physical structure, and L(H), the mathematical structure, it seems that L(E) contains abstract entities—namely 'hypothetical' quantum events—which are not nominalistically acceptable. Our representation function simply maps one platonistic structure to another. How can we get a nominalistic analogue of L(E)?



In his 1996 paper "Towards a Nominalization of Quantum Mechanics," Mark Balaguer puts forth the following thesis: the closed subspaces of H can be taken as representing *physically real propensity properties*—in particular, the r-strengthed propensity of a Ψ-state system to yield a value in Δ for a measurement of A. Recall that states can be thought of functions from (A,Δ) to r in [0,1]. So each state specifies a set of pairs $<(A,\Delta), r>$ (one pair for each event in S(E)), each of which in turn determines an r-strengthed propensity property—denoted by (A,Δ,r)—of a system to yield a value in Δ for a measurement of A. Out of the set of these propensities, S(P), we can then construct a lattice L(P) which is homomorphic to L(H), similarly to our method for S(E). Unlike L(E), though, L(P) has no abstract yet-to-occur quantum events; instead it has physically real propensity properties—actual and current dispositions to behave in a certain way. We can then complete the program by giving nominalistic versions of the lattice-theoretic predicates, analogous to the metaphysically primitive betweenness and congruence relations that Field introduced in his nominalization of distances in spacetime (Balaguer 1996, "Towards…" pg. 217-8).

[9] A partially ordered set is an ordered pair <A, R> where A is a non-empty set and R is a reflexive, transitive and asymmetric relation defined on A. Orthomodular lattices are a special type of partially ordered set with certain properties (e.g. has a minimum and maximum element), as explained in (Balaguer 1996, "Towards…" pg. 216).
[10] In fact they are *iso*morphic.



3. OBJECTIONS TO BALAGUER'S NOMINALIZATION ATTEMPT

While the nominalization attempt outlined above certainly seems more successful than the version Malament struck down (assuming we are entertaining the idea that propensities, unlike hypothetical quantum events, can be physically real), it does have weaknesses and missing parts. One missing part is, as Balaguer admits, the dynamics:

> "…what is left unnominalized is the dynamics of the theory—in particular, the Schrödinger equation. But I don't see any reason why this can't be nominalized in the same general way that Field nominalizes the differential equations of Newtonian Gravitational Theory… I do not know of any arguments against the nominalizability of the dynamics of QM" (Balaguer 1996, "Towards…" pg. 223).

I will present some such arguments. First however, I wish to discuss Otávio Bueno's evaluation of Balaguer's nominalization in his 2003 paper, "Is It Possible to Nominalize Quantum Mechanics?" This discussion will provide context for my own views, and clarify some important points in the debate. Bueno raises three main objections: (i) Balaguer's strategy is incompatible with other interpretations of QM, (ii) we cannot nominalize the relation that compares propensity strengths, and (iii) propensities are abstract, just like events. I will explore each in turn.

Regarding (i), Balaguer argues that his program does *not* assume a propensity interpretation of QM. He says that he is only committed to the broad claim that "quantum probability statements are about physically real propensities of quantum systems," which, he says, can be understood in a very *weak* way as saying that that "quantum systems are irreducibly probabilistic" (Balaguer 1996, "Towards…" pg. 217). Bueno points out that there are anti-realist interpretations, for example van Fraassen's modal interpretation,[11] that

[11] This interpretation rejects Balaguer's assumption that Observable B has value *b* if and only if a measurement of B is certain to have outcome *b*.

would endorse the latter claim but deny that this entails the former. Moreover, Hidden Variable Theories, like Bohmian mechanics, are incompatible with both claims. Bueno then presents the following dilemma:

> "If incompatible with a whole family of interpretations of QM, then [Balaguer's] strategy is inadequate, since it is not capturing the underdetermination of interpretations typical of QM… [And if compatible] there won't be nominalistically acceptable replacements for quantum structures for Balaguer's strategy to succeed… In either case, the strategy does not seem to go through. The upshot is that we should not expect to settle the issue about which interpretation of QM is adequate based on whether nominalism in the philosophy of mathematics is true!" (pg. 1434).

This analysis strikes me as uncharitable. Of course, one nominalization strategy will not be compatible with a "whole family" of QM interpretations. Compare two non-orthodox QM interpretations: Ghirardi-Rimini-Weber theory (GRW), a spontaneous collapse interpretation, and Bohmian mechanics, a deterministic hidden variable theory in which particles have definite positions and are guided by a 'wavefunction field.' The ontologies and dynamics of these two theories are *drastically different*, as are parts of their mathematical structures. I imagine the nominalization of Bohmian mechanics could take a form similar to that of NGT, whereas the nominalization of GRW may have the same broad Fieldian ideas behind it, but (if the last three pages are any indication) all the details would be completely different—in fact I suspect they would need to be different from the last three pages as well.[12] But it is precisely because of the fundamental differences between the interpretations of QM that we should *not* place Bueno's burden on the nominalist. The nominalist may try to nominalize the various interpretations of QM with different strategies, and then it will be up to progress in philosophy of QM to decide which interpretation to take on board. Trying to nominalize specific interpretations is *not* automatically equivalent to trying to settle which interpretation is adequate based on this effort (and this is not what Balaguer was doing). While Bueno would be right in saying that the nominalist e.g. Balaguer should be more *cautious* in clarifying which interpretation(s) his strategy applies to, it seems unfair to use the underdetermination of interpretations

[12] This is because we would have to remove reference to measurement and instead assign to each particle a fixed propensity to spontaneously collapse. We would have to replace talk of measurement yielding values in Δ with talk of collapse yielding a Gaussian distribution of values in Δ. So the whole definition of $CS(A,\Delta)$ would have to be adjusted; it is unclear whether we could end up getting GRW analogues of L(P) and L(H) that are homomorphic. Thus Balaguer's strategy may even be incompatible with this popular collapse theory.

as grounds to place that strategy in such an unreasonable dilemma. Therefore, I will proceed with a discussion of Balaguer's approach, and take the reader to understand that these details, in addition to my central thesis as a whole, may not apply in the context of other QM interpretations.

Objection (ii) concerns whether Balaguer can nominalize the probability assignment *r*; since *r* is a real number in [0,1], and numbers are not nominalistically acceptable. Balaguer does this by introducing a predicate–the "propensity between" relation—which compares propensities between different physical entities (e.g. electrons) in different states:

> "Thus we replace [platonistic] sentences like, 'State-Ψ electrons have *r*-strengthed propensities to yield values in Δ for measurements of *A*', with [nominalistic] sentences like, 'State-Ψ electrons are *(A,Δ)-propensity-between* state-Ψ_1 electrons and state-Ψ_2 electrons' " (Balaguer 1996, "Towards…" pg. 225).

This primitive betweenness relation is supposed to be directly analogous to a betweenness relation for points in space that we could use to nominalize distance. However, there is an important difference: for the point-betweenness relation, we need only commit to the existence of a point *y* spatially located on the line segment whose endpoints are *x* and *z* (Bueno pg. 1430); and these points are *already spatially ordered in physical reality.* In the probability case, what determines the ordering of the propensities? We cannot use the ordering from the probability assignments *r* themselves, since those numbers do not exist for the nominalist. Luckily, Δ is a Borel set that will give the correct probability ordering—hence the strange name for the relation, "*(A,Δ)-propensity-between*," used in the quotation. But, as Bueno points out, Δ is a set of real numbers, and thus a Platonic object—so we cannot use *it,* either! Unlike the distance case, where we do not need to "move beyond" the points to obtain the betweenness relation, the *(A,Δ)-propensity-betweenness* relation is tied to Δ (1430). Thus, it seems that even if propensities are physically real, we cannot nominalize the lattice-theoretic predicates without taking the probability assignments in [0,1], or Δ, to exist; at least one of these entities is indispensable. I will return to this issue.

Objection (iii) challenges Balaguer's claim that propensities are not abstract. One reason we might think a particular physical property is not abstract is that it is causally efficacious. But, as Bueno argues, "*an infinite collection of propensities is not something causally efficacious.* A propensity is only a particular *disposition* to behave in a given way; it is not something that in

itself has already happened, or has been actualized" (1430). Another reason we might think a physical property is not abstract is that it is located in spacetime—and dispositions indeed are. But, as Bueno points out, "[t]he problem is that we cannot simply stop here. The disposition or propensity presupposes a *modal* component" (1431). So claims like the following have to be true: if I *were* to send that electron through this certain kind of magnet, it *would* move upwards (in this example, r is 1). But why is this true? Its truth cannot be the result of what goes on in a possible world, since possible worlds are abstract.[13] It seems that what was nominalistically unacceptable about events—that some have not yet occurred—also applies, in an important sense, to propensities.

I will now raise the worry that I alluded to—call it (iv)—regarding the dynamics of QM. I suspect Balaguer was right, in the earlier quote, that a nominalization of the Schrödinger equation (SE) itself would not encounter too many problems. What concerns me, however, are the points in time during which SE *stops* describing the evolution of the system—namely, *the points in time at which measurement(s) of observable(s) on the system occur*. These non-linear interruptions to the 'flow' of SE are precisely the moments when this talk of propensities ought to become dynamically relevant, since we know the following two empirical claims, **A** and **B**, are true (I use measurements of electron spin in different directions as an example):

> **A**. If we take an ensemble of electrons, which are spin-up in the z-direction, and measure them for spin in the x-direction, then approximately 50% will be spin-up (since $r = 0.5$).
> **B**. If we take a single electron which is spin-up in the z-direction, and measure it for spin in the x-direction, and then keep re-preparing the system and repeating the measurement, then after a sufficient number of trials we will find that the percentage that are spin-up approaches approximately 50% (again since $r = 0.5$).

It seems, at least *prima facie*, that the platonist can account for both of these facts quite easily: the probability assignment for spin-up is 0.5 and we know from the Law of Large Numbers (LLN) that the average of the results obtained from a large number of identical trials should be close to expected value, and should approach the expected value as we perform more and more trials (in **A**, the number of trials is the number of electrons in the ensemble, and in **B** it is the

[13] But what if the possible worlds are taken to be *concrete*? Bueno says that, since such worlds are not *actual*, the nominalist cannot use talk of them to support the truth of modal claims (1431). However, Bueno does not give a detailed justification for this conclusion. It may be interesting to consider this objection in the context of a nominalization of the Many Worlds interpretation of QM.

number of repetitions). It follows from LLN that the associated frequencies will converge to the objective probability of 0.5. But how can the nominalist account for these facts? How can she give an account of what is occurring physically which does not presume the *existence* of expected values or probabilities as numbers, and the *truth*—not just logical consistency—of LLN as a mathematical theorem? Note that LLN makes use not only of the expected value but also of the fact that the *number* of trials *n* becomes large: $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$ $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$. If (ii) is correct, then we cannot even give an ordering of the propensities which admits that they are halfway *(A,Δ)-propensity-between* electrons in the *spin-down-in-x* state (*r=0*) and electrons in the *spin-up-in-x* state (*r*=1). And even if we overcome (ii) and deny (iii) in favor of the position that propensities *are* nominalistically acceptable, then we are still left with the following problem: if Bueno is correct that propensities are not causally efficacious, then we cannot say that the physical propensity properties of the electrons are *causing* things to obtain such that approximately half the outcomes are spin-up. So why does the result (the associated frequency of ~0.5) obtain *at all*?

To deal with (iii) and (iv), it may be tempting to make the following adjustment to Balaguer's thesis: let us claim that particles have physically real propensity properties and that these propensities *are* causally efficacious. How? As time flows and measurements occur, they cause outcomes to occur with frequencies associated approximately with the strength of the propensity toward those outcomes. Note some peculiar features of this kind of "causality." For single-trial cases, the propensity has no clear causal role with regard to determining the outcome (we might want to say the propensities do have some sort of role: they make it more likely for certain outcomes to occur—but here the very point of this propensity interpretation of chance is to use talk of physically real propensity properties to *replace* talk of probabilities and likelihoods). For cases like **A**, where the trial electrons are separated spatially (perhaps lightyears apart), the propensities act *collectively* and *non-locally* to ensure that approximately half the outcomes are spin-up. For cases like **B**, where the trials are separated temporally, the electron's propensity acts *through time* to ensure approximately half the outcomes are spin-up. We can then imagine combining **A** and **B** into a situation of repeated measurements of an ensemble, in which case the electrons' propensities must act collectively and through time to continue to ensure the ~0.5 associated frequency. Finally, to deal with (ii), we add that it is a brute fact about physical reality that the propensities are ordered by strength; we discover the continuum of relative strengths empirically by noting the relative frequencies with which different outcomes occur.

Of course, this adjusted thesis is nowhere near robust (and I have not

even begun to try to nominalize it—for example, the associated frequencies are still left as numerical ratios). We may wonder whether it can give a remotely plausible account of why, when we increase the number of electrons in **A** or number of repetitions in **B** (or the number of electrons *and* number of repetitions in the combined situation), we get an associated frequency that is closer and closer to the propensity strength (i.e. the crux of LLN). Moreover, we may question whether this story of "collective action" on the behalf the propensities of each electron in the ensemble of **A** is really possible if the propensities are supposed to be properties of *individual* particles (making it seem like **A** should be an example of multiple simultaneous single-trial cases, where I argued propensities have no real causal role). We may therefore need to stop talking of propensities as properties of individual particles, and rather as properties of some overall system.

4. A DEEPER WORRY FOR THE FICTIONALIST

Suppose that, due to a combination of the difficulties above, we are unable to nominalize QM (and, more specifically, the role of chance in the dynamics of QM). Say, for example, that probability assignments, Δ, and LLN are indispensable to the theory. What can the Fictionalist say in response? In another 1996 paper, Balaguer makes the compelling point:

> "If there exist any mathematical objects, then [according to PCI] they are not causally relevant to the physical world (and QM doesn't entail otherwise); thus, the behavior of the physical world will be the same whether or not there exist mathematical objects (and QM doesn't entail otherwise); thus, QM's picture of the physical world will have the same degree of accuracy whether or not there exist mathematical objects" ("A Fictionalist…" 305).

The upshot is that, while the Fictionalist may be unable to nominalize QM, she *should* be able to 'list off'—perhaps in an unattractive and jumbled way—all the 'nominalistic content' of QM (all the facts about what is happening on the physical level, expressed in nominalist language) and still end up with a *complete* picture of the quantum world. Of course, this picture will not be theoretically attractive (so we haven't really nominalized a *theory*) but it will still contain enough information to constitute a full description of physical reality. Otherwise, it would seem to suggest that the existence of Δ or the truth of LLN could affect physical reality—but this contradicts PCI! Once the Fictionalist has this complete picture, she can claim that we only need to commit to the truth of the nominalistic content

of QM, since "the [dispensable and indispensable] fictional mathematical claims are not *part* of what's being said about the physical world; they are, rather, part of the apparatus which enables us to say what's being said" (Balaguer 1996, "A Fictionalist…" pg. 307).

So how exactly should the Fictionalist 'list off' the facts that we *need* the mathematical apparatus to express? Balaguer shows how this can be done with one canonical fact, and claims that we need "do nothing more than to run through the theory and do what I just did for this one fact." What he says is worth quoting in full (the italics are mine):

> "For instance, one such fact… is that if we take an ensemble of electrons which are spin-up in the z-direction and measure them for spin in the x-direction, then half of them will be spin-up. (Actually, it won't always be the case that *exactly* half are spin-up, just as it's not always the case that *exactly* half the tosses of a fair coin are heads. *But this problem can be solved by speaking in terms of relative frequencies or propensities; thus, for instance, we might say that the larger the ensemble of z+ electrons being measured for [spin in the x-direction], the closer we will get to the result that for every electron which is measured spin-up, there corresponds a unique electron which is measured spin-down, and vice-versa)…*all the mathematical baggage here… is doing nothing but providing a convenient and precise way of describing purely nominalistic facts about the quantum world" (Balaguer 1996, "A Fictionalist…" pg. 301).

However, as I argue in (iv), this problem is *not* solved by speaking in terms of propensities unless we also give a nominalist account of, first, the relationship between propensities and associated frequencies (for example, by arguing as I tried to earlier that propensities somehow *cause* associated frequencies), and, second, how increasing the size of the ensemble should bring us closer to the result indicated by the propensity strength. The problem is also not solved by speaking *just* in terms of relative frequencies, since then we would run into the severe problems associated with frequentist interpretations of chance.[14] Indeed, it seems that to express facts like **A** and **B** in nominalistic terms—and thus to capture the complete picture of the dynamics of the quantum world via purely nominalistic content, we *do* need to nominalize the role of chance.

[14] Frequentism conflicts with our intuition that a fair die has a probability of 1/6 of landing on one of its faces, even if we only throw it once, and even if we throw it seven times but find that it always lands on the same face. Hypothetical frequentism (that we take the frequency given an infinite number of identical trials) will not work here, since we will most likely need to use modal operators that Bueno objects to in (iii).

5. GENERALIZING THE WORRY

I argue that the Platonist *also* has a stake in this nominalization. Recall **A**. The objective probability that the measurement will yield spin-up is 0.5 (for the moment we are forgetting about Balaguer's propensity thesis). A natural view is that this chance *explains* the associated frequency of ~0.5 we observe. But it is also a key part of the concept of chance that *it is unlikely for associated frequencies to diverge from chances*. So it is unlikely after, say, 1000 trials that we would find only ~0.1 of the electrons to be spin-up in x. What explains this unlikelihood? Perhaps we can simply answer: it is unlikely because LLN says it is, and LLN is *true*. But recall that LLN only tells us that *there is a low chance* that the expected value and the average empirical value will diverge. And we cannot chalk up our explanation to a low *chance*, since then we have just pushed the issue back one step further: what explains the low frequency with which *that* low chance diverges from *its* associated frequency (the low frequency at which we observe larger divergences than LLN predicts in the spin experiment)? I propose that to explain this relationship between chance and associated frequency we may need to bestow chance with a sort of weak, one-directional causal role in our metaphysics, like what I did for propensities: the role of causing associated frequencies to obtain. As with the propensity case, this kind of 'causation' would be strange; it would have to be non-local and occur 'through time,' as it were.

However, now the Platonist faces an ontological question: are these chances abstract mathematical objects, or are they physical properties? If they are mathematical objects, this would entail a violation of PCI, since I just posited that chances interact causally with the physical world. And if they are indeed purely physical properties, then we should be able to give a complete description of what is physically going on in cases like **A** and **B** by simply listing off all the nominalistic content of QM—after all, if Platonic objects are causally isolated, in principle they should not affect physical reality. Therefore, nominalistic content should be sufficient in capturing all the brute physical facts about that reality. But, as I argued before, expressing facts **A** and **B** in a nominalistic way is not going to be easy without a proper nominalization of the role of chance.

Indeed, it seems one reason the question "what is the relationship between chance and associated frequencies?" is so difficult to answer (and remains such a central puzzle in the metaphysics of chance) is that the nominalization program does not easily extend to it. In particular, it appears difficult to describe what is happening without not only implying that mathematical objects/laws exist, but also implying that they play a sort of causal role. I propose a corollary to this analysis, which is more relevant to the

metaphysics of chance than to the Platonist versus Anti-Platonist debate: the nominalist perspective may be a useful framework within which to approach the above question—if not because nominalization techniques will actually help lead us to the answers, then because they will at least help us keep in mind the separation between our mathematical and physical structures. And, as long as we are all still endorsing PCI, admitting this separation will help narrow down the number of metaphysical accounts we may consistently adopt.

6. CONCLUSION

I have shown that, contrary to Balaguer's suspicion, there are problems that face the nominalization of the dynamics of QM, particularly in accounting for the frequencies associated with measurement outcomes. Without a nominalization of the role of chance in these situations, the Anti-Platonist will not only be unable to nominalize QM, but also unable to 'list off' the canonical QM facts Balaguer has in mind. And, given the constraints of PCI, this inability affects the Platonist too. The upshot is that both parties must nominalize the role of chance. Indeed, *anyone* who wishes to provide a complete picture of the dynamics of the quantum world without turning abstract mathematical objects/laws into causal agents must meet this challenge—regardless of her interest in the nominalist program. It is worth pointing out that these issues lie at the intersection of highly unresolved debates in the philosophy of math, QM and the metaphysics of chance. They will be difficult to discuss without using controversial assumptions and arguments from all three areas. However, this should not stop us from trying.

BIBLIOGRAPHY

Balaguer, Mark. A Fictionalist Account of the Indispensable Applications of Mathematics. *Philosophical Studies*. 83: 291-314. 1996

Balaguer, Mark. Towards a Nominalization of Quantum Mechanics. *Mind*. 105: 209-26. 1996

Bueno, Otávio. Is It Possible to Nominalize Quantum Mechanics?. *Philosophy of Science*. 70: 1424-36. 2003

Field, Hartry. *Science Without Numbers*. Princeton: Princeton University Press. 1980

Malament, David. Book Review: Science Without Numbers. *Journal of Philosophy*. 79: 523-34. 1982

On the Reason to Reduce the Efficacy of Moral Luck

Dallas M. Ducar
The University of Virginia

Moral luck applies to any circumstance where an agent is clearly not in full control of an action, despite being assigned either moral blame or praise. Most ethical theories agree that if an agent performs an action voluntarily and is without outside coercion, the agent is to be held responsible for the action. However it appears that an agent, while appearing to be in full control of both her actions and the consequences of said actions, may rarely or never have full control of both. Thus, the problem of moral luck presents itself. An action and its consequences appear to be affected not only by volition but also by environmental and genetic factors as well. Moral luck, therefore, is composed of environmental and genetic factors. In these factors rests constitutive luck, luck in who one is, or in the traits and dispositions that one has.[1] More importantly, constitutive luck includes both contingent (e.g., inclinations, capacities and temperament) as well as the necessary features of a person that are beyond the person's control.[2] Imagine that Smith and Jones are both are driving down a highway in separate trucks. Now imagine in both scenarios a little girl walks into the middle of the highway and both Smith and Jones slam on their breaks as quickly as possible. Jones is able to move quickly and dexterously to ensure his truck stops before hitting the girl. However, Smith (dealing with the same environmental factors as Jones) is unable to manipulate his vehicle fast enough, due to a slower reaction time, and unfortunately rips the little girl to bloody shreds. Smith will feel guilt, pain, and possible suffering from legal implications, while Jones will suffer from little to no repercussions. This is the problem of constitutive luck, that factors such as our genetics can influence and even cause individuals to commit morally reprehensible actions outside of personal agency.

However, due to recent and continuing developments in the field of reproductive technology, it appears that it is indeed possible for humans to alter moral luck. If this is indeed a possibility, then there may be good reason to reduce the influence of moral luck whenever possible. For example, say it is possible to isolate and alter genes related to dexterity for Smith prior to birth. If the altering of such constitutive genes were possible and resultantly, Smith's dexterity was enhanced, the probability of unjustified moral blame lessens. This is not to say that moral luck becomes non-existent. Yet the probability of harm resulting from genetic factors outside Smith's agency will decrease if he is not subject to genes which can alter his volitional capabilities. If we can increase the amount of moral responsibility an agent is capable of we can then judge the agent based on her own decisions rather than dispositions.

Before discussing how one is to limit constitutive luck, it is useful to understand two main distinctions in genetic intervention. One can differentiate

[1] Nagel, Thomas. *Mortal Questions*. Cambridge: Cambridge University Press. 1979 (pg. 28)
[2] Lippert-Rasmussen, Kasper. Justice and Bad Luck. *The Stanford Encyclopedia of Philosophy*. Fall 2009.

between genetic enhancement and therapy (also known as negative and positive genetic engineering). Allen Buchanan describes this as a defined boundary in bioethics. "In many contemporary discussions, the negative/positive distinction is used to draw a fundamental moral boundary," describes Buchanan. "There is a presumption that negative genetic interventions—are morally permissible, whereas positive interventions are morally impermissible or at least highly problematic."[3] For the purposes of this essay I will use the contemporary nomenclature of genetic therapy and enhancement. The goal of genetic therapy is to treat or prevent a disease, while enhancement concerns itself with the augmentation of a capability or attribute. While the distinction is necessary, both actions have the ability to lessen the efficacy of moral luck. However, permissibility of genetic therapy, I think, would be less morally contentious. Instead, I will focus on illustrating that we have a good reason to use genetic enhancement to lessen the efficacy of moral luck, provided that they augment universally desirable traits. Before continuing, I will first show the moral permissibility of both therapy and enhancement.

One can limit the efficacy of constitutive luck via enhancement. Even if we agree some enhancement may be morally permissible, there may still remain worries regarding agency. Enhancement appears to be more contentious due to the fact that if an agent is being externally enhanced, the agent is not consenting to the enhancement itself. The reason to limit the effect of moral luck does seem dubious if we are infringing upon agency. Thus if we desire to limit the effect of constitutive luck, there must be an attempt to change the constitutive self without infringing upon agency. Similar to constitutive luck, the constitutive self includes both contingent (e.g., inclinations, capacities and temperament) as well as the necessary features of a person.[4] Therefore, if one enhances dexterity for a potential individual such as Smith, there appears to be little choice in his own constitutive self. Any external enhancement for Smith will not result in him choosing his own traits and dispositions. Instead, there must be an approach to allow the hypothetical Smith to have a "voice" in his constitutive self ex-ante. However, before continuing with this approach it is important to note the possible harms of enhancement.

Harms to the child must be taken into primary consideration when considering genetic interventions. The worry is that enhancement will allow for person affecting harm. According to a person affecting view of harm, "a person is harmed by an act if she is made worse off than she would otherwise have been if that act had not been performed."[5] Philosophers such as Julian Savulescu argue that genetic selection is able to navigate around this worry due to the fact that it

---

[3] Buchanan, Allen; Brock, Dan; Daniels, Norman; and Wikler, Dan. *From Chance to Choice: Genetics and Justice*. New York: Cambridge University Press. 2000 (pg. 106)

[4] Lippert-Rasmussen, Kasper, Justice and Bad Luck.

[5] Savulescu, Julian; Hemsley, Melanie; Newson, Ainsley; and Foddy, Bennett. Behavioural Genetics: Why Eugenic Selection Is Preferable to Enhancement. *Journal of Applied Philosophy*. 23(2): 157–71. 2006 (pg. 162)

is not person affecting. Savulescu claims, "When we change somebody's body in a way they dislike, unless they have given consent, it counts as a harm."[6] If Savulescu is correct, then genetic enhancement could possibly never be morally permissible in limiting the efficacy of moral luck. This is because one would be promoting constitutive luck, only this time it would be guided by external desires instead of luck. Instead, Savulescu argues for genetic selection. "When considering a possibly-enhancing intervention, there are two possible futures for the same person," notes Savulescu. "If an individual is born from genetic selection, persons with and without the predisposition to genetic disorders would be different people."[7] This different-persons view shows that there will consequently be no harm in selection because we can select one embryo against another. One embryo would cease to exist; however, there would be no resultant harm to the embryo. On the other hand, the embryo that was selected would be expected to continue growing and the harm would be person affecting. The key notion which Savulescu posits is that genetic selection will not result in person affecting harm.

However, it is genetic enhancement, not genetic selection that will be unable to lessen the efficacy of moral luck. While genetic selection concerns itself with different number choices, enhancement is concerned with same number choices. Before continuing, I will define both cases, originally defined by Derek Parfit. Same person choices are those in which the same person will result regardless of how one acts, while different person choices result in different people existing. Within the realm of these choices are same and different number choices. A same number choice results when you have the same number of people, while different number choices are what is implied. The result of the different number choice is a different amount of people than originally.[8] The key is that selection does not result in the same number while enhancement does. This is because of the very contention that Savulescu raises; enhancement is person affecting. This is important because if we wish to enhance rational agency for an individual there must be only one organism that we are concerned with. It is only person affecting choices that will allow us to change the grapple of moral luck on an individual.

The problem in relation to moral luck is that a non-person affecting choice will not result in any change of constitutive luck towards a set individual. As mentioned before, this is due to the fact that the individual will never change in a constitutive self if there is no person affecting choice. Rather, it is enhancement which will result in the altering of the individual to alter the constitutive self. Thus, while there may be the existence of a person affecting harm, enhancement

---

[6] Ibid., pg. 162
[7] Ibid., pg. 163
[8] Parfit, Derek. *Reasons and Persons.* Oxford: Clarendon Press. 1984 (pg. 154)

remains the only option to increase rational agency and lessen the effect of moral luck. We must then find a method to ensure that these person affecting harms are no longer valid. In particular it must be illustrated that these person affecting worries are indeed not harms. If the action is intended not to harm the individual, there must then some way to show that the individual would consent and that the action is indeed beneficial to the individual. I must show that enhancement can be permissible if I wish to argue for a good reason to limit moral luck by extending rational agency via enhancement.

In understanding why some enhancements are permissible, one must look to John Rawls' idea of primary goods. In *A Theory of Justice* Rawls outlines a list of primary goods which he says are, "things which a rational man wants whatever else he wants." Primary goods are things which every rational person should value, regardless of their concept of the good life. Rights, liberties, opportunities, income, are just some of Rawls' primary goods which are listed. [9] These are assumed to be goods which every agent would desire and to decline these goods would be irrational. The universally desirable traits I promote are not these social goods, however they are like them. Universally desirable traits are those which every rational person should value, similar to primary goods. In relation to enhancement, universally desirable traits are those which every rational being would desire (E.G., strength, eyesight, intelligence). I will first posit that enhancement is morally permissible if it and only if it serves to augment universally desirable traits. This is because these moral enhancements are only allowed by means of consent, even if it is hypothetical. Thus, if a rational agent is to be born and we can enhance the agent via universally desirable traits, it is presumed that the agent would consent.

In the case of enhancement, the problem of controlling one's constitutive luck can be solved through universally desirable traits. As stated before, universally desirable traits are those which every rational person should value and thus all rational agents would consent to the maximization of universally desirable traits. This model of universally desirable traits allows rational consent to exist without the need for an agent to *physically* consent to a change in constitutive self. Therefore, Smith would be enhanced to allow for the maximization of universally desirable traits so that the efficacy of moral luck could be limited. If Smith's dexterity was enhanced so that he was able to make the quick movement of pulling on the truck's breaks before it hit the pedestrian, there would be no question of wrongfully assigning blame. Enhancing this capability directly translates into extending rational agency which can lessen the impact of moral luck.

Having shown moral permissibility and the need for universally

[9] Rawls, John. *A Theory of Justice*. Cambridge: Harvard University Press. 1999 (pg. 54)

desirable traits, I will now show why there is a good reason to augment Rawlsian universally desirable traits through enhancement. If we are provided with genetic information of how to enhance an individual to augment universally desirable traits, there will be greater chance that moral praise and blame can be correctly assigned. This reason is a deontological one, based on the notion of promoting rational agency and thereby limiting the effect of moral luck. A universe with more rational agency is more desirable than a universe with less; this is at least a prima facie claim. An agent that has greater rational agency is responsible for more elements of an action and its consequences. Thus, a universe with this increased moral responsibility is one where agents can more accurately be judged without as much influence from moral luck. This will then allow us to more accurately assign blame or praise to a rational agent. If one is to accept a good reason to promote rational agency, then one must desire to promote a universe with more accurate praise or blame. I suspect that this idea is a non-contentious one. Therefore, this good reason to limit the efficacy of moral luck follows from the promotion of rational agency. I posit that genetic enhancement is one method of limiting the efficacy of moral luck, if and only if it augments universally desirable traits. Thus, there would be good reason to promote such talents and abilities such as: strength, speed, sight, intelligence, creativity and all other faculties that fall under universally desirable traits.

Before continuing, I would like to make the necessary distinction between maximization and promotion of universally desirable traits. Opponents may claim that engaging in enhancement will allow for limitless enhancement or maximization for the sake of itself. This is known as the slippery slope argument. One may argue that when limiting obstacles for persons we may be blurring the line between enhancement for promotion of rational agency and enhancement for maximizing all capabilities. My argument does not invoke any hedonistic utilitarian calculus; instead it simply concerns itself with the notion of universal agreement.

Proponents of a consequential outlook tend to elevate procreative autonomy quite highly. Julian Savulescu argues for the idea that we have a moral obligation to select for the best children, also known as procreative beneficence. This argument inevitably posits the selection of not only disease genes, but non-disease genes, even if this maintains or increases social inequality. “Couples (or single reproducers) should select the child, of the possible children they could have, who is expected to have the best life,” Savulescu argues. “Or at least as good a life as the others, based on the relevant, available information.” Both procreative beneficence and the augmentation of universally desirable traits argue for promoting a good; however, universally desirable traits are normative while procreative beneficence is dependent on subjective notions such as procreative

autonomy. Savulescu posits that the welfare of the child is of primary consideration and yet he allows for the selection of non-disease genes. Therefore, procreative beneficence allows for the possibility to change non-universally desirable traits such as eye color or height. It is the necessity of enhancing solely for universally desirable traits that distinguishes the limiting of moral luck from the promotion of Savulescu's procreative beneficence and maximization.

Engaging in consequentialist worries of what will limit the efficacy of moral luck could result in the promotion of the aforementioned slippery slope argument. For example, imagine a couple that decide that their offspring must be enhanced to be tall or else he could be subject to constitutive luck. The couple could easily imagine a situation which illustrates the necessity of a person being tall for moral blame to be appropriately assigned. They could imagine a situation where their potential child could rescue a cat from a tree or any other hypothetical. The problem with this is that one could just as easily imagine a situation where being tall would be a detriment. This is why the limitation of moral luck must be grounded in normative rather than subjective claims. Any individual can imagine a thought experiment where one may be able to justify a trait in order to limit the efficacy of moral luck. Instead, it is necessary to only enhance for goods which every agent would desire and to decline the goods would be irrational.

The necessity of everyone agreeing to the enhancement of a trait allows for a sufficientarian claim. I am not claiming that limitless enhancement is a good everyone desires. This is hardly the case. Instead, in the case of another example such as sight, it does not seem suspect that everyone would agree with enhancement to allow for slightly increased vision. However, it seems dubious that every agent would agree with increasing eyesight to see infrared and ultraviolet naturally. Universally desirable traits allow it to be possible to decline such a good because it can be quite problematic at times. It is not difficult to imagine a situation in which infrared vision could be plain annoying and result in not being a desirable trait for everyone. Thus the need for universally desirable traits is quite evident to escape the worry of the slippery slope objection.

Even if there is no maximization for the sake of itself, there appears to be another worry. There is a striking argument known as the deaf culture argument which argues that even though being deaf limits some opportunities, the deaf culture provides valuable opportunities and benefits to its members. If it is true that these opportunities are being limited, then one could argue that agency is being limited in this scenario. This would inevitably lead to the idea that hearing is not a universally desired trait. However Buchanan shows deafness may actually limit the consequential opportunities of the individual. "Even if it could be shown that the distinctive benefits of sign language are only available to the deaf, it is one thing to say that those who are deaf gain a great good from

this mode of communication. It is much less plausible to say that a reasonable person confronted with a choice between suffering the limitations of deafness while gaining the benefit of this mode of expression and avoiding the limitations of deafness but not being able fully to appreciate the unique expressive power of sign language would choose the latter."[10] Buchanan shows that it is quite plausible to believe that no rational being would desire deafness for the strict reason to join a deaf community.

The decision to use genetic intervention to produce a deaf child appears quite dubious. Deafness is not a trait which every rational person should value, regardless of their concept of the good life. Additionally, Buchanan's previous argument appears to hold in showing why hearing would be a universally desirable trait. The enhancement of this trait, hearing, would allow for more agency and therefore the lessening of moral luck. Also hearing can allow one to be more responsible for their actions due to the furthering of capabilities. If possible, the decision of intervention is much better presented to "a reasonable person confronted with a choice *ex ante*."[11] The individual, in this case, is much better off making the decision for herself, whether she would like to join the deaf culture or not.

However, some would still argue against the notion of enhancement. Michael Sandel argues for a notion of giftedness, meaning an attribute that is not deserved but still given, and above "normalcy." In particular, Sandel argues that, "Genetic enhancements undermine our humanity by threatening our capacity to act freely, to succeed by our own efforts, and to consider ourselves responsible—worthy of praise or blame—for the things we do and for the way we are."[12] This suggests a diminished moral agency of the person who has undergone enhancement. In particular, Sandel argues for a giftedness of life, which is to "recognize that out talents and powers are not wholly our own doing, despite the effort we expend to develop and to exercise them."[13] Therefore constitutive luck, according to Sandel, should be valued rather than limited. He claims there is a certain virtue in our inequalities and striving for excellence, rather than simply inheriting excellence. While I do not argue for excellence but rather an augmentation of universally desirable traits, Sandel is opposed to much of enhancement.

Additionally, Sandel claims that parents must be "open to the unbidden" and not attribute so much to choice. The gradual drift from chance to choice, according to Sandel, allows "Parents to become responsible for choosing or failing

---

[10] Buchanan, Allen; Brock, Dan; Daniels, Norman; and Wikler, Dan. *From Chance to Choice: Genetics and Justice*. (pg .282)
[11] Ibid.
[12] Sandel, Michael. *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Cambridge: Harvard University Press. 2007 (pg. 74)
[13] Ibid., pg. 78

to choose." Sandel is concerned with a form of hyperagency that results in him arguing against varying notions of enhancement. Sandel shows this by explaining the notions of accepting love and transforming love. He claims that the two balance each other out; one accepts the child while the other seeks the well-being of the child. Too much of one form of love creates a vice, and Sandel posits that there is too much transformative love to the point where perfection is sought. However, this argument against choice in the realm of enhancement is directly opposed to the good reason that parents must undertake to extend the will and limit moral luck. Both of Sandel's arguments rest on a certain value of constitutive luck.

I shall respond to Sandel's arguments by employing some ideas of Frances Kamm. One notion which Kamm particularly focuses on in response to Sandel is the idea of nature's giftedness. As Kamm describes, "we treat when we eliminate a dysfunction, not merely present anything that interferes with nature's gifts. Dysfunction is an interference with healthy human life."[14] We can still appreciate the giftedness and still supplement it with something new. According to Kamm, we can still express our appreciation of giftedness while allowing for genetic intervention. He also posits that fixing dysfunction is different than enhancement because "it alone has a virtue of accepting the normal and avoiding the implied rejection of normal human life."[15] However, there still must be room for enhancement to go beyond the norm. This is because an augmentation of universally desirable traits can result in more capabilities for the individual and thus more agency. This agency allows the individual to be more responsible for praise or blame without being limited to their constitutive self. As previously noted, to limit moral luck one must go beyond treatment.

Sandel argues that our transformative love is overtaking accepting love and yet Kamm argues that enhancement does not show a lack of accepting love. Kamm categorizes changes made before the child exists as *ex-ante* changes and those made once a child exists as *ex-post* changes. The argument is then made that before the existence of a person, there is no person we have to lovingly accept. He claims, "Not accepting whatever characteristics nature will bring but altering them ex-ante does not show lack of love."[16] Sandel does not illustrate clearly that pursuing enhancement for children ex-ante is inconsistent with a proper balance between accepting and transforming love. Rather, it seems equally permissible to claim that enhancement ex-ante can be done out of love for the child and the parents willingness to limit the effects of moral luck. Accepting love can only be employed *ex-post*. Therefore, if one can invoke the use of universally

[14] Kamm, Frances. What is and is Not Wrong with Enhancement? *American Journal of Bioethics*. 5(3): 100-36. 2006 (pg. 106)
[15] Ibid., pg. 107
[16] Ibid., pg. 113

desirable traits ex-ante to enhance goods that every human being would want, it is very difficult to see how love is non-existent.

Another worry is that enhancement will change our notion of humanity. These worries of depriving humans of experience are posited by philosophers such as Erik Parens. In particular, he asks whether our attempts at enhancement will detract from important facts of human experience and inadvertently impoverish us. Parens is concerned particularly with human "fragility" and claims that enhancement may, "reduce the change and chance to which we—creatures whose forms are largely determined by the genetic hand dealt us by nature—have hitherto been subject."[17] The notion of chance is important for Parens; he attributes much of our appreciation of humanity to the extraordinary combination of human effort and chance. There is thus a certain value which Parens places on human endeavor and triumph. Thus, in our case of moral luck, there is an appreciation of the constitutive luck which is endowed via the genetic lottery. Parens would not, I think, object to enhancing a child to have a slightly enhanced eyesight if possible (given that it was universally accepted). However, he does allow for the possibility that removing certain obstacles to humanity could impoverish human experience.

Parens finds value in certain obstacles, which may be considered elements of constitutive luck. Therefore, Parens may find enhancement anywhere past the "garden variety" morally suspect. The questionable nature of creating a more "perfect world" with fewer obstacles goes directly against the idea of limiting moral luck. Parens argues that this too could alter our relationship with nature. Yet, our relationship with nature has changed for millennia. Proponents of enhancement such as Fritz Allhoff claim that as time progresses our *standards of evaluation* change as well. "Consider, for example, the Olympians of classical Greece," he writes. "Their athletic accomplishments, although tremendous at the time, could be duplicated now by even the most average inter-collegiate athlete."[18] This is possible due to advancements in sports medicine, training, nutrition, wellness and more. "Excellence and accomplishment is measured relative to some standard, and that standard is dynamic,"[19] Allhoff claims. Therefore, the comparative advantage, as he calls it, does not impede achievement, but instead affects the standard by which achievement is measured. Using the aforementioned example, our standards are fluid and thus even the most basic major league baseball player could be seen as enhanced by the ancient Olympians.

Allhoff's argument for dynamic standards of evaluation illustrates that

---

[17] Parens, Erik. The Goodness of Fragility: On the Prospect of Genetic Technologies Aimed at the Enhancement of Human Capacities. *Kennedy Institute of Ethics Journal*. 5(2): 141-53. 1995 (pg. 143)

[18] Allhoff, Fritz. Germ-Line Genetic Enhancement and Rawlsian Primary Goods. *Kennedy Institute of Ethics Journal*. 15(1): 39-56. 2005 (pg. 51)

[19] Ibid.

advantages do not necessarily challenge the notion of achievement. For the case of moral luck, one must also realize that there are dynamic standards developed through society and technology. Attributes which may have been considered subject to chance years ago can now be considered universally desirable traits. A serf in medieval Europe may have been considered an "unfortunate" when born with what was considered normal vision in his time period. However, enhancement will allow him to no longer be subject to nearly as much constitutive luck as he would be. If an individual can be enhanced via universally desirable traits, it is difficult to see how the individual becomes impoverished. This is especially important because a universally desired trait would most likely not be considered extreme at the time. Instead, it is much more plausible that universally desired traits would allow for more moderate enhancement that would seemingly not detract from this notion of human experience. It seems plausible that enhancing eyesight can allow for a more fulfilled and abundant life. Parens does not deny that this can be a possibility, instead he asks us to; "think more deeply about how attempts at control and alteration that truly enhance life are different from those that impoverish it."[20] This argument of contemplation is not in direct opposition to the notion of universally desirable traits; however, a deep analysis of enhancement is still necessary before any decisive action is taken.

It is clear that moral luck is a problem especially in the realm of ethics; however I have that its efficacy can be lessened through genetic interventions. Interventions which allow for more capabilities will then lead to more rational agency which, I have shown, is capable to limit moral luck. However interventions ex-ante, such as enhancement, are morally permissible if and only if they serve to augment universally desirable traits. Enhancement must be restricted to universally desirable traits as it created a form of hypothetical consent to the enhancement of one's own constitutive self. This hypothetical consent must be comprised of decisions that every rational person should value, regardless of one's conception of the good life. There is therefore good reason to promote rational agency whenever possible to allow for an increased moral responsibility. This will then allow us to more accurately assign blame or praise to a rational agent. It is the combination of enhancement and universally desirable traits that can allow individuals to flourish and create a universe with increased moral responsibility and decrease the efficacy of constitutive luck. Genetic technologies have no longer made us subject to moral luck. Instead, we have the capability to alter moral luck and good reason to do so.

[20] Parens, Erik. The Goodness of Fragility: On the Prospect of Genetic Technologies Aimed at the Enhancement of Human Capacities. (pg. 150)

BIBLIOGRAPHY

Allhoff, Fritz. Germ-Line Genetic Enhancement and Rawlsian Primary Goods. *Kennedy Institute of Ethics Journal.* 15(1): 39-56. 2005

Buchanan, Allen; Brock, Dan; Daniels, Norman; and Wikler, Dan. *From Chance to Choice: Genetics and Justice.* New York: Cambridge University Press. 2000

Kamm, Frances. What is and is Not Wrong with Enhancement? *American Journal of Bioethics.* 5(3): 100-36. 2006

Lippert-Rasmussen, Kasper. Justice and Bad Luck. *The Stanford Encyclopedia of Philosophy.* ed. Edward N. Zalta. Fall 2009

Nagel, Thomas. *Mortal Questions.* Cambridge: Cambridge University Press. 1979

Parens, Erik. The Goodness of Fragility: On the Prospect of Genetic Technologies Aimed at the Enhancement of Human Capacities. *Kennedy Institute of Ethics Journal.* 5(2): 141-53. 1995

Parfit, Derek. *Reasons and Persons.* Oxford: Clarendon Press. 1984

Rawls, John. *A Theory of Justice.* Cambridge: Harvard University Press. 1999

Sandel, Michael. *The Case Against Perfection: Ethics in the Age of Genetic Engineering.* Cambridge: Harvard University Press. 2007

Savulescu, Julian; Hemsley, Melanie; Newson, Ainsley; and Foddy, Bennett. Behavioural Genetics: Why Eugenic Selection Is Preferable to Enhancement. *Journal of Applied Philosophy.* 23(2): 157–71. 2006

Wily Socrates:

How Seriously Should We Take the Arguments of Hippias Minor?

Michael Allen Ziegler
The University of Virginia

I. AUTHENTICITY

Before I proceed, I will take this opportunity to defend the authenticity of the Hippias Minor. As John M. Cooper notes in his introduction to the dialogue in *Plato: Complete Works*, some commentators, fearing for the moral reputation of Socrates because of the dialogue's final argument (which seems to make out the man who does injustice voluntarily to be better than the one who does so involuntarily), have deemed the dialogue to be spurious.[1] While I will address this specific concern below, I believe that there is good evidence to suggest that the dialogue is authentic, largely because it is cited by Aristotle in his Metaphysics[2]. Granted, the citation does not mention Plato as the author[3], and it has been suggested to me in conversation that in fact it might be the case that the dialogue could have been written by another member of the Academy. Fair enough, but the question of authorship is difficult for any ancient text, and I tend to agree with Cooper that Aristotle cites works of Plato in other places without the use of his name, presuming that his audience would be familiar with the author of these works. In fact, we should be more inclined to believe that the Hippias Minor is the work of Plato because of Aristotle's failure to mention the author of the work, as when this occurs in his writings he is usually referring to a work of Plato.[4] I think, then, that we ought to be skeptical about claims that the dialogue is spurious.

II. WILY SOCRATES

On the whole, the Hippias Minor deals with the topic of falsehood. How appropriate is it, then, that Socrates appears at his wiliest in this dialogue? As it begins, Eudicus speaks to Socrates after the conclusion of a speech given by Hippias on the topic of the Iliad, alluded to by Hippias in the Hippias Major[5]. He notes that Socrates has been silent throughout the performance and now afterwards, is neither giving praise nor cross-examining any of Hippias' points.[6] Socrates replies that, now that he mentions it, Eudicus' father, Apemantus, used

---

1 Smith, Nicholas D., trans. "Lesser Hippias", in *Plato: Complete Works*. ed. John M. Cooper and D.S. Hutchinson. 922-936. Indianapolis: Hackett. 1997 (pg. 922)

2 Ross, W.D., trans. "Metaphysics", in *The Complete Works of Aristotle*. ed. Jonathan Barnes. 1552-1728. Princeton: Princeton University Press. 1984 (1025 $a^{2\text{-}11^3}$)

3 Ross' translation does mention Plato by name, but this is an editorial decision on his part.

4 For a discussion of this, see Jowett's Appendix 1 to his translation of Hippias Minor. Though Jowett tends to think poorly of HM's quality of literary style, he takes Aristotle's citation as important evidence for the dialogue's authenticity, citing a case where the Phaedo, a dialogue whose authenticity is not in dispute, is cited by Aristotle without specific mention of Plato as author.

5 Woodruff, Paul, trans. "Greater Hippias", in *Plato: Complete Works*. ed. John M. Cooper and D.S. Hutchinson. 898-921. Indianapolis: Hackett. 1997 (286 $b^{4\text{-}5}$)

6 Smith, trans. "Lesser Hipp.", pg. 923 (363 $a^{1\text{-}4}$)

to say that the Iliad was better than the Odyssey insofar as Achilles was a better hero than Odysseus.[7] This exchange sets up the intellectual pretext for the dialogue between Socrates and Hippias—for Socrates to ask Hippias about Achilles and Odysseus—but it should be noted that Socrates' silence suggests that his intentions may go beyond his intellectual interests. Later in the dialogue Socrates describes how he always seeks to understand a speaker better if he thinks that he has said something wise, but, "If the speaker seems to me to be someone worthless, I do not ask questions, nor do I care what he says."[8] Given the fact that Socrates was silent following Hippias' speech, it seems to follow that he considers Hippias to be a "worthless person" (phaulos). We might question what intent Socrates has in questioning Hippias if he has such a low opinion of the man. Surely this is done in part to please Eudicus since he is the one who urges Socrates to question Hippias in the first place, but one might wonder if Socrates' motives do not involve something other than gaining greater understanding given the fact that he does not seem to consider Hippias to be a worthy contender.

Taking into account his poor regard for Hippias, the flattery Socrates bestows on him begins to look rather disingenuous. Socrates goes so far as to say that the fame of Hippias "… is a monument for wisdom to the city of Elis and to your parents" (364 b). He may think Hippias is a fool, but Socrates goes so far as to call Hippias a treasure for his hometown. It should be noted that Hippias does not seem to mind this flattery; in fact, he eagerly accepts it. For example, after Socrates asks him if he is experienced in geometry, Hippias responds, "Egoge": "I am indeed." Hippias is quite ready to accept any praise, direct or indirect, that Socrates gives him, all the time not realizing that Socrates thinks him "phaulos." To understand what motivation Socrates has for this duplicitous flattery, and indeed his motivation for pursuing an elenchus of someone he considers worthless, I believe we must look to the overall dramatic structure of the dialogue, which I believe can be construed as a sort of speaking contest between a wily Socrates and a dull Hippias.

III. THE SPEAKER'S CONTEST

Why ought we to think of the Hippias Minor as a speaker's contest between Socrates and Hippias? It certainly seems that within the Platonic

[7] Ibid., 923 (363 b$^{2-4}$)
[8] Ibid., 928 (369 d$^{3-4}$)

corpus as a whole, Socrates is wary of speeches.[9] In fact, at 373 a[3-4], Socrates warns Hippias that giving a long speech will not aid in his coming to a clearer understanding of what they have been discussing.[10]

It is important to recall, however, that Hippias does suggest a speaker's contest after the two have reached an impasse about whether the truthful person and the liar are the same person. He suggests that each man give a speech arguing for which man (Achilles or Odysseus) is better, "and these men (houtoi) will know which of us speaks better."[11] Here Hippias is suggesting that the dialogue revert to an activity he is quite comfortable with: giving a long speech in a competition and having an audience decide who has spoken better. As Hippias mentions before, he has often gone to Olympia to compete, and for as long as he has been going he has never encountered anyone that has been better than him at anything.[12] It is because he always answers anyone asking about what he has prepared for speaking that it would be shameful for him to avoid the questioning of Socrates. For Hippias, then, there is something inherently competitive about public speaking, and the conversation he and Socrates have been having up to this point has taken place in front of an audience: the few remaining behind after his speech, including Eudicus. Perhaps sensing that Socrates has been getting the best of him so far in the eyes of the audience, Hippias attempts to shift to a style of competing about which he feels a "godlike state of mind," as Socrates describes it.[13]

While Hippias might want to shift the structure of the "competitive" dialogue he and Socrates have been having, Socrates does not seem to take up the challenge, ignoring the offer. However, Socrates then proceeds to give a speech, on the basis of evidence from the Iliad, which makes Achilles seem as wily as Odysseus.[14] How wily of Socrates, to give a speech quite similar to the one Hippias suggested that he give, while seeming not to accept Hippias' challenge. For good measure, Socrates gives a second, relatively long speech[15], whose tone is really not so different from some of the boasting that Socrates says he has heard from Hippias in the agora[16], with Socrates speaking about his desire to inquire further of people he thinks have said something wise and calling this his one virtue. Perhaps the content is not so different from what

---

[9] A notable exception to this can be found in Republic Book 1 at 348 a-b. Socrates gives Thrasymachus the option of either taking turns giving speeches and having those speeches judged by the audience or engaging in a dialogue so that they can act as "jury and advocates." It is clear that this style of debate is familiar to Socrates, though he expresses a dislike for its use both in the Hippias Minor (373 a) and in the Protagoras.

[10] Smith, "Lesser Hip.", pg. 932 (373 a[3-4])

[11] Ibid., 929 (369 c[5-6])

[12] Ibid., 923 (364 a[6-7])

[13] Ibid., 923 (364 a[1])

[14] Ibid., 929-930 (369 d-370 e)

[15] Ibid., 931-932 (372 a-373 a)

[16] Ibid., 928 (368 b-369 a)

we find in his description of his divine mission in the Apology, to seek out those who seem to have wisdom and question them, but the tone seems more boastful in this speech. Socrates is not simply trying to advance his divine mission of seeking out those who seem to be wise and, hopefully, gaining knowledge from them, but he is also beating Hippias at his own game, giving a couple of subtle and boastful speeches in front of an audience.

While Eudicus might not seem to have much of a role in the dialogue, if we view the dialogue as a sort of speaking contest, then we can see Eudicus as playing the role of a spokesperson for the audience and also a referee for the implicit contest between Socrates and Hippias. After Socrates gives his second long speech, we might get the impression that Hippias has had enough and is trying to slink off, and this is when Socrates says, "…I might justly call for your help, too, son of Apemantus, for you goaded me into a discussion with Hippias."[17] As Eudicus was the one who persuaded Socrates to question Hippias and then asked Hippias if he would talk to Socrates, he is uniquely positioned to keep the conversation going. He reminds Hippias that, "…he wouldn't flee from any man's questioning."[18] And why is this the case? Because whenever in the past he has gone to the festival of the Greeks at Olympia, he answers whatever anyone asks him of things he has prepared for display.[19] To bow out at this point would be akin to losing a speaker's contest at the Olympics, and as Hippias himself says before, "Ever since I began taking part in contests at the Olympic games, I have never met anyone superior to me in anything."[20] How unfortunate would it be for Hippias to now be completely bested by Socrates in speech-making, all while an audience looks on? As a representative of this audience, and the one who initiated the "contest" in the first place, Eudicus is well placed to keep Hippias talking to Socrates. Unfortunately for Hippias, however, he comes off as the far duller of the two, unable to answer to Socrates' arguments and failing to press any of the salient objections that he does make. To explore this topic further, we now turn to evaluating some of the arguments of the Hippias Minor and see where several of them might be vulnerable to attack.

IV. EVALUATING THE ARGUMENTS

Having dealt largely with the literary component of the text, we now turn to three of the arguments offered in the Hippias Minor and see where,

[17] Ibid., 932 (373 a^{6-7})
[18] Ibid., 932 (373 b^{2-3})
[19] Ibid., 923 (363 c-d)
[20] Ibid., 923 (364 a^{6-7})

if anywhere, Socrates goes wrong in his arguments. It is my contention that while these arguments are not fallacious, strictly speaking, they are certainly misleading and two of them are quite vulnerable to attack while the third has serious moral consequences but a limited degree of plausibility. Given what has been said before about the wily character of Socrates in this dialogue, this might not come off as all that surprising.

Socrates establishes that when Hippias says that Odysseus is wily, he is saying that he is a liar.[21] From this initial conclusion, and what Hippias tells him in the course of his elenchus, Socrates comes to the conclusion that the liar and the truthful man are not opposites, as Hippias has said before, but are in fact very much the same. I believe we can begin addressing the flaws in this argument by understanding better the meaning of the words Socrates uses in the original text for "liar" and "to lie."[22] The Ancient Greek words pseudes and pseudomai, most properly mean "false" and "to speak false things," respectively. The concept of deceiving, to intentionally deceive someone as regards the truth of the matter, can be captured in Greek by pseudomai, but it can equally mean "to speak false things," without any intention to deceive. When Socrates speaks of a liar in the Greek, he uses the adjective "pseudes" as a substantive, which has the stricter meaning of a "liar" when applied to persons[23]. However, Socrates takes care to differentiate between voluntary (hekon) liars and involuntary (akon) liars. A young girl who gives an incorrect answer to a math problem could be called "pseudes," not because she was trying to deceive her classmates, but because she spoke something that was false. She would be an "involuntary liar," as she does speak false things, but does not do so from intent to deceive. Along with Cooper's note on the term "liar" in Hippias Minor, I tend to read pseudes in the text as "liar," but also, more broadly, "one who speaks false things."[24] It is my belief that Socrates relies on a degree of ambiguity between these two different senses of "pseudomai," lying and speaking false things, when he says that Achilles, and others, are pseudes.

We have a clear case of the sort of problems with the argument that arise from this ambiguity around 367a. Previous to this, Socrates has gotten Hippias to admit that liars are in fact knowledgeable and powerful when it comes to lying. To be powerful in lying, Socrates has pointed out, is to be able to lie when the liar wants to lie. Taking the example of arithmetic, he asks Hippias if he could give the answer to a math problem and Hippias says that of course he could.

---

21 Ibid., 924 (365 $b^{4\text{-}5}$)

22 My source for the Greek words used in the text comes from a course packet that is a photocopy of George Smith's 1895 edition of Platonis Ion et Hippias Minor for the Upper Forms of School. While this edition of the text is currently out of print, it can be found online on the website of the University of Toronto Libraries.

23 See entry on pseudes in LSJ.

24 Smith, "Lesser Hip.", pg. 924

Then Socrates makes the crucial move to bring together the liar and the truthful person, as he asks Hippias, "Could you lie the best, always consistently say falsehoods about these things, if you wished to lie and never tell the truth?"[25] Socrates suggests that the person who does not know the truth about the matter would not be able to do so: "Don't you think the ignorant person would often involuntarily tell the truth when he wished to say falsehoods, if it so happened, because he didn't know…"[26] I think it's fair to say we would all agree with Socrates that having the knowledge necessary to accomplish some task, such as giving the correct answer to an arithmetic problem, also implies the ability to do the opposite, such as giving the wrong answer to an arithmetic problem. In fact, ex hypothesi, the person who has knowledge of arithmetic is the only person who has the power to give incorrect answers to an arithmetic problem whenever she wants, as the person who lacks all knowledge of arithmetic will not know whether she is giving correct or incorrect answers. Where I differ with Socrates (and where I think Hippias ought to) is that this does not imply that the person who lacks the knowledge of a particular subject required to give the correct answer cannot be deceptive about that subject, and, particularly in the case of arithmetic, be nearly as effective a liar as the person who has knowledge of arithmetic. Let us return to the example of the child who does not know the answer to an arithmetic problem and has no knowledge of arithmetic. She does not have the power to give the correct answer to the arithmetic problem when she wants and does not have the power to give the incorrect answer when she wants; she is essentially giving random answers precisely because she does not know which answers are correct, and which are incorrect. However, what makes her so miserable is the fact that she consistently gets the answer wrong, since she is ignorant of arithmetic. It is true that the student who is knowledgeable in arithmetic would be able to most consistently give wrong answers to arithmetic problems if she wants, and so perhaps is "better" at giving a false answer than the child who does not have knowledge of arithmetic, but it seems to me, in the example of arithmetic, they would both be pseudes about as much as they wanted to be, since the girl who does not have knowledge of arithmetic would have a difficult time stumbling onto the right answer and the girl who knows the right answer could always avoid it. They would both be "pseudes" pretty consistently. So when Socrates rhetorically asks, "Then who becomes a liar about calculations, Hippias, other than the good person?"[27] (i.e. the person good at calculation), I must disagree with him. Strictly speaking, the child who does not have knowledge of arithmetic is not really able to give an incorrect answer

[25] Ibid., 926 (367 a^{1-2})
[26] Ibid., 926 (367 a^{3-4})
[27] Ibid., 927 (367 c^{4})

when she wants, but if some schoolmate of hers whom she dislikes asks her for help with an arithmetic problem, she will likely give whatever answer comes to mind, and this answer will most likely be incorrect. Perhaps it is not correct to say that she is deceiving her schoolmate about arithmetic per se, but instead about her knowledge of arithmetic. If she gives an answer to her schoolmate that happens to be incorrect, then in the broadest sense of the Greek word she is pseudes, having spoken a false answer. She is also pseudes in the sense of being a liar, having deceived her schoolmate into believing that she has given her a correct answer, when in fact she does not know if the answer is correct or not. She might stumble upon the right answer, but that would be fairly difficult to accomplish. I grant Socrates that the person who is truthful, the person who has the knowledge to give the correct answer, is also most capable of giving a false answer, but this does not imply that someone who does not have the knowledge of a given subject cannot speak false things about that subject or, in a certain way, be deceptive about that subject so as to become a liar.

This ambiguity of "pseudes" and "pseudomai" also comes into play when Socrates makes arguments to the effect that Achilles is just as good as Odysseus insofar as he is as good a liar. Socrates cites a number of passages from the Iliad in which Achilles speaks of sailing home, yet nowhere does he seem to prepare to go home.[28] In addition, Socrates points to the differing answers Achilles gives to Odysseus and Ajax in Book 9 of the Iliad, telling the former that nothing will persuade from sailing for Phthia and saying to the latter that he will not fight until Hector reaches his tent. Socrates argues that Achilles must be a bold and talented liar in order to get away with contradicting himself in front of Odysseus to get the better of him.[29] Now, in the grand scheme of things, we might say that the statements Achilles makes about sailing home are false since Achilles never sails home, and so Achilles might be called pseudes in the broadest sense of the word, in that he does speak false things. Is it also the case that he is pseudes in the sense of being a voluntary liar? Hippias protests that Achilles is not a voluntary liar about going home, since he is "forced to stay and help by the misfortune of the army."[30] I take Hippias to be claiming that Achilles is not a liar, but that he says false things as a result of a change in intention. He is not voluntarily deceiving the Greeks as to his intentions, but his change in intention does render false the statement that he intends to sail home to Phthia the next day. Achilles gives seemingly contradictory responses to Odysseus and Ajax because "his good-naturedness (euetheia) led

[28] Ibid., 930 (370 d^{3-4})
[29] Ibid., 930 (371 d^{5-6})
[30] Ibid., 930 (370 e^{5-6})

him to say something different to Ajax and Odysseus."[31] That Hippias does not suggest that Achilles lies here, either voluntarily or involuntarily, indicates that he sees this as another case of a change of intention. Hippias contrasts this with Odysseus, who only lies or tells the truth from some malevolent plotting. Socrates replies by asking "Didn't it emerge just now that voluntary liars are better than involuntary ones?"[32] Quite a complicated question! How should we understand "better" here? As above, we can certainly grant Socrates that those who are able to lie voluntarily, such as the person who has knowledge of arithmetic, are better (if at times, only marginally) at deceiving than those who lie involuntarily, who speak false things with the intent of deceiving without knowing that the things they are speaking are false. However, applying this distinction to Achilles is rather difficult if we think, as Socrates suggests, that he must be a good liar, and so, a voluntary one. Of course, it is quite possible for someone to lie about her intentions voluntarily, but could she really do so involuntarily, without knowledge of her intentions? Leaving aside questions of subconscious intentions, it seems implausible to me that a person could involuntarily lie about her intentions, that is, to speak false things about her intentions while not knowing what her intentions are. A person's intentions may change over time, but then she is not lying when she speaks of having an intention that runs contrary to her previous intention. There is no "better" liar in the case of lying about one's intentions since there is only one kind of liar, the voluntary one. So then I think we ought to be very sympathetic to Hippias when he objects that Achilles is not lying. He may be speaking "false things," but that would only be because his intentions have changed since he first raged against Agamemnon. Socrates is relying upon the ambiguity of "pseudes" and "pseudomai," as we might say that Achilles is pseudes in the broadest sense of the word, but would have a more difficult time applying the term to him the in the narrower sense of being a liar.

Perhaps the most problematic argument, and the one that I believe Socrates himself takes most seriously, is the argument at the end of the dialogue, which seems to lead to the conclusion that the one who does injustice voluntarily is none other than the good man. Socrates' argument goes something like this: Let us assume that justice is a sort of knowledge or ability or both. If this is the case, the more knowledgeable or able a person is, the greater their capacity for justice. The person who would have such a knowledge or ability would appear to be the just man. "To refrain from injustice is to do something fine," whereas to do something unjust is to do something shameful, according to Socrates.[33]

[31] Ibid., 931 (371 e^{1-2}
[32] Ibid., 931 (371 e^{7-8})
[33] Ibid., 935 (376 a^{1-2})

Although Socrates does not mention it explicitly, it seems to follow that the man who has the knowledge of justice seems to be the one who is able to accomplish fine things voluntarily and know they are fine things. Presumably, he will wish to accomplish fine things. However, just as Socrates said before, to have a power or ability of some kind also seems to imply the ability to do the opposite. If this is the case, the man who does injustice voluntarily would seem to be the just man, since he is the only one who has the ability to commit injustice voluntarily. The man who does not have knowledge of what is just will not be able to commit injustice voluntarily because he does not know what it is to commit injustice, though he may go around committing many unjust acts without realizing that they were unjust. The problematic conclusion of the argument is that the good man, who will possess a soul that is good and able to do justice, is also the only one who will do injustice voluntarily[34] (376b). It appears that such a man would have knowledge of what is fine and have the ability to do what is fine, but instead would voluntarily do what is shameful, according to what Socrates has said before. Unlike other arguments in the dialogue, this one has teeth, just so long as the condition, "if there is such a person,"[35] is satisfied. Common sense seems to indicate that there are people who do injustice voluntarily, as Hippias suggested when he said that the law treats these people more harshly than those who do injustice involuntarily. However, the only people who can do injustice voluntarily are those who have knowledge of what is just, and these people are the ones we call just, according to the argument. So it seems that the law is in fact punishing the just man if it punishes those who do injustice voluntarily. However, it should be noted that we must fulfill the condition, "if there is such a person." It might be the case that even though the just man has the capacity to commit injustice voluntarily, there is no such person who is both just and also voluntarily does injustice, and indeed I think both the moral psychology found in the so-called "Socratic" dialogues and that found in the Republic do forbid such a case.[36] Putting this aside for the moment, let us consider the attitude of Socrates in the dialogue towards his own arguments.

V. HOW SERIOUSLY DOES SOCRATES TAKE HIS OWN ARGUMENTS?

The above question is the impetus for this paper, but even given what

[34] Ibid., 936 (376 b$^{5\text{-}7}$)

[35] Ibid., 936 (376 b^{7})

[36] The moral psychology of the Republic does suggest that a man who has knowledge of the just might be overcome by the appetites to commit injustice in a case of so-called "clear-eyed" acrasia, but of course the lack of harmony between the parts of the soul would also seem to preclude this man from being truly just. On the Socratic intellectualist account, the just man will never act against his knowledge of what is just.

I have argued so far, I do not mean to suggest that the answer I propose is anywhere near definitive. On the basis of what has been laid out in the sections above, I will argue that the only argument Socrates takes slightly seriously is the argument about the just man at the end of the dialogue, and that we should be more skeptical of just how seriously Socrates takes the conclusions of the two earlier arguments.

We established above that Socrates is being fairly disingenuous in engaging in dialogue with Hippias; while the prompting of Eudicus might act as a pretext for his commencing the conversation, it seems to me that the persistence of Socrates, even after Hippias proves so inept at following his arguments much of the time, cannot be attributed to a desire to gain greater understanding, but a desire to beat Hippias at his own game. Hippias boasts at the beginning of the dialogue about his prowess in speaking, and all Socrates does is take what Hippias says and draw a conclusion from it that makes Hippias' statement that the liar and the truthful man are entirely opposites self-contradictory. As I have argued above, I do not think that the first two of Socrates' arguments stand up to scrutiny, but it is not as if he does not give Hippias plenty of opportunities to scrutinize them. He begins with the example of arithmetic, then goes onto geometry, then astronomy, and then describes all the knowledge and crafts Hippias possesses.[37] Hippias, however, is not quite up to the task of figuring out where Socrates goes wrong, as he says that Socrates always picks at the most difficult part of the argument but neglects to elaborate as to how that picking is unfair.[38] On such a foolish opponent, why not use arguments that are less effective than the best you have, just to twist him in knots with? It is the equivalent of a successful sports team playing down to the level of competition of an inferior team, doing well enough to win comfortably, but not blow them out of the water. Similarly, after Hippias suggests a speaker's contest between the men, Socrates gives a speech on the basis of evidence, but his argument is also slightly off. Hippias puts up his most spirited defense at this point, but Socrates keeps up the assault. At this point, Hippias makes a very cogent objection, arguing that the law treats those who do injustice involuntarily less harshly than those who do injustice voluntarily.[39] Unfortunately for him, Socrates is not playing fair and addresses his objection by saying that, "it appears entirely the opposite."[40] Socrates then asks Hippias to cure his soul so that he no longer sees those who do evil voluntarily as better than those who do so involuntarily. But is that not what Hippias was attempting to do with the example of the law?

---

[37] Ibid., 926-927 (366 c-368 a)
[38] Ibid., 929 (369 c^{1-2})
[39] Ibid., 931 (372 a^{4-5})
[40] Ibid., 931 (372 d^{4-5})

If Socrates' objective in putting forward these arguments was to obtain greater understanding, he has a funny way of going about it, not engaging Hippias when he gives a serious objection. My contention is that he is not interested in greater understanding, but only in making Hippias look like a fool since he considers him to be worthless anyway. So if his argument about Achilles is faulty and Hippias offers an interesting objection, it would make sense for him to ignore it; he is not interested in correcting a faulty argument, just so long as Hippias is given a hard time.

Additionally, we must keep in mind that the dialogue takes place in front of an audience, and on the basis of this I believe that we may draw a couple more important conclusions that cast doubt on the seriousness with which Socrates takes his own arguments.

Eudicus indicates at the beginning of the dialogue that after the conclusion of Hippias' speech, "we who have most claim to have a share of the practice of philosophy are left to ourselves."[41] It is clear that the audience observing the dialogue is composed of the sort of young men who follow Socrates around, spending time around him to try and engage in the practice of philosophy. Socrates does not just engage in the dialogue with Hippias for the sake of Eudicus, it seems, but perhaps also to provide his audience of eager, young philosophy students with arguments to examine for themselves. Perhaps the arithmetic argument is not airtight, but it does provide the audience with a chance to consider the fact that the ability to accomplish something, such as to give a correct answer to an arithmetic problem, also implies the ability to do the opposite, such as to give the incorrect answer to an arithmetic problem. When Socrates glosses over Hippias' objection that the law treats those who do injustice voluntarily more harshly than those who do injustice involuntarily at 372d, we might imagine that he does so in part to give his audience the opportunity to consider what Hippias has said and, like the readers, wonder why Socrates would not dignify such a sensible suggestion with a response. We might see the dialogue with Hippias, rather like the written dialogue itself, as a pedagogical exercise for Socrates' philosophy-minded audience.

I have argued above that Socrates attempts to shame Hippias in Hippias Minor, and I believe that the inclusion of an audience, and how in this the dialogue differs from the Hippias Major, might give us a clue as to why Socrates attempts to do this. Unlike the Hippias Minor, the dialogue of the Hippias Major appears to take place in the context of a private conversation, and I believe that we might see the addition of an audience in the former dialogue as an occasion for shaming Hippias for the betterment of his soul, along similar lines to that

[41] Ibid., 923 (363 a[5])

suggested by Brickhouse and Smith in their book *Socratic Moral Psychology*. It is my belief that we should see the Hippias Major and Hippias Minor as being dramatically connected, with the action of the Hippias Major occurring a few days before, as well as alluding to, the speech Hippias has just made before the beginning of the Hippias Minor.[42] In the course of the Hippias Major, Hippias continually fails to answer and even comprehend Socrates when he asks him to tell him about what "the fine itself" is. Instead, he thickly continues to give examples of fine things. In fact, at the end of the dialogue, Hippias seems to be no less aware of ignorance of what "the fine itself is" than when he started speaking to Socrates, and even implicitly insults him by saying that the "friend" of Socrates who wished to know what the fine itself is should stop "applying himself… to babbling nonsense."[43] Socrates attempts to engage intellectually with Hippias in the Hippias Major, and this fails miserably. Having realized just how thickheaded Hippias is, Socrates does not attempt to engage intellectually with him in the Hippias Minor, but instead decides to shame him in front of an audience, beating him at his own game of engaging in a speaker's contest in front of an audience. Why would Socrates want to publicly shame Hippias? In Socratic Moral Psychology, Thomas Brickhouse and Nicholas Smith suggest that the theory of moral psychology found in the "early" or "Socratic" dialogues includes a conative role played by the appetites and passions, such as pride and shame. They cite a prominent passage from the Apology where Socrates appears to say that using shame is part of conducting his divine "mission."[44] In this passage Socrates imagines himself speaking to one of the people of Athens, saying to him, "Good sir, since you're an Athenian, aren't you ashamed about having as much money… as you can and you don't care about… making your own soul as good as possible?"[45] (29e) While I do not intend for this paper to constitute an in-depth endorsement of Brickhouse and Smith's theory of a Socratic moral psychology, I do believe we might see the actions of Socrates in the Hippias Minor as characteristic of an attempt to better Hippias' soul by shaming him. Having failed to make Hippias realize the extent of his ignorance in the Hippias Major, Socrates resorts to making him feel ashamed by making him look foolish in front of an audience in the Hippias Minor. Whereas we find an ignorant Hippias continuing to exalt the virtues of speechmaking at the end

---

[42] At 286b in the Hippias Major, Hippias says that he will be giving a speech about Neoptolemus and Nestor "the day after tomorrow," and that "Eudicus, Apemantus' son, invited [him]," the very same Eudicus whom we find in the Hippias Minor.

[43] Woodruff, "Greater Hippias", pg. 921 (304 b^{6-7})

[44] Brickhouse, Thomas C. and Smith, Nicholas D. *Socratic Moral Psychology*. New York: Cambridge University Press. 2010 (pg. 57)

[45] Grube, G.M.A., trans. "Apology", in *Plato: Complete Works*. ed. John M. Cooper and D.S. Hutchinson. 17-36. Indianapolis: Hackett. 1997 (29 d^{6-8}-e^{1-3})

of Hippias Major[46], we see Hippias noticeably less confident in Hippias Minor, apparently trying to sneak off towards the end of the dialogue before Eudicus asks him to stay.[47] (373a) For him to attempt to leave the conversation indicates that he has experienced quite a change in demeanor from the beginning when he boasted that "…since I began taking part in the Olympic games, I have never met anyone superior to me in anything."[48] I think we should be sympathetic to Hippias when he says that "….Socrates… creates confusion in arguments, and seems to argue unfairly,"[49] since Socrates has argued rather unfairly, but it seems that Hippias does not feel so confident in his ability to combat and overcome the unfair argumentation of Socrates, and perhaps fears that his failings will reflect shamefully upon him. For Hippias to even seem to fail to beat an opponent in a speaking contest would certainly wound his pride, and it is the audience that adds these higher stakes to this conversation.

As for the argument at the end of the dialogue, that the man who commits injustice willingly is the just man, I believe that Socrates really does take this argument seriously, and that we should take him at his word when he says that he "waver[s] back and forth and never believe[s] the same thing."[50] It should be noted that he does say that he wavers back forth, "as I said before, on these matters,"[51] and this refers back to 372d, where Socrates uses the same word (planomai) to describe his vacillation about the fact that "those who harm people and commit injustice and lie and cheat and go wrong voluntarily, rather than involuntarily, are better than those who do so involuntarily."[52] This argument is truly a troubling one to Socrates, and he wavers back and forth on it as if he were afflicted with a disease. It is related to the earlier arguments in its form, but unlike the examples of arithmetic and the intentions of Achilles, this argument has some dire moral consequences. If even Hippias, whom Socrates calls wise (though obviously does not consider to be wise), and other wise men are not able to find a solution to this problem, then we common people are really in trouble. If the law treats more harshly those who voluntarily do injustice, and the ones who voluntarily do injustice are the just men, then, on the face of it, we seem to have a problem with our laws, since they will be condemning just men. As I have said, I think that Plato's solution to the problem is to deny that there is any man who has knowledge of what is just and also voluntarily does injustice; he simply will not do injustice. While Socrates is troubled by this argument,

---

[46] Woodruff, "Greater Hippias", pg. 921 (304 b^{6-7})
[47] Smith, "Lesser Hippias", pg. 932 (373 a^{6-7})
[48] Ibid., 923 (364 a^{7-8})
[49] Ibid., 932 (373 b^{4-5})
[50] Ibid., 936 (376 c^{3})
[51] Ibid., 936 (376 c^{2-3})
[52] Ibid., 931 (372 d^{5-8})

he points to this solution by making his conclusion contingent on the fact that "there is such a person."(376b)[53] Indeed, the fact that he wavers back and forth on the argument suggests a lack of serious commitment, as terribly troubling as the argument's conclusion might be.

VI. LITERARY PURPOSE

Why does Plato put arguments in the mouth of Socrates that do not stand up to much scrutiny? My tentative answer is that it is to engage the reader in the dialogue and the arguments being put forth. Even if Socrates argues unfairly at times, as Hippias suggests, then Hippias should do better at pointing out where Socrates has gone astray, and correct him. Like Hippias, we, as readers, may be perplexed by some of the moves Socrates makes, but then it is our place to step in and diagnose if and where they go wrong. Like the dialogue's audience, which Eudicus informs us is interested in philosophy, the reader should evaluate the seemingly cogent counter-example of the law treating those who do injustice voluntarily more harshly than those who do so involuntarily and judge for herself whether or not she believes that Socrates should give such short shrift to it. Like Socrates, the reader may struggle with the idea that the just man is the only one who is able to do injustice voluntarily, but hopefully this will engender further dialogue about the subject in the mind of the reader. We have good reason to doubt Socrates' commitment to some of the conclusions he reaches in the Hippias Minor, but this does not mean that the arguments do not have merit as subjects of debate for students of philosophy, whether they consist of the dramatic audience in the dialogue or the literary audience reading the dialogue.

[53] Ibid., 936 (376 b[9])

BIBLIOGRAPHY

Brickhouse, Thomas C. and Smith, Nicholas D. *Socratic Moral Psychology*. New York: Cambridge University Press. 2010

Grube, G.M.A., trans. "Apology", in *Plato: Complete Works*. ed. John M. Cooper and D.S. Hutchinson. 17-36. Indianapolis: Hackett. 1997

Grube, G.M.A., trans. "Republic", in *Plato: Complete Works*. ed. John M. Cooper and D.S. Hutchinson. 971-1223. Indianapolis: Hackett. 1997

Jowett, Benjamin. Appendix I to Lesser Hippias. Gutenberg Project. Published electronically: http://www.gutenberg.org/files/1673/1673-h/1673-h.htm#link2H_INTR. 2008

Ross, W.D., trans. "Metaphysics", in *The Complete Works of Aristotle*. ed. Jonathan Barnes. 1552-1728. Princeton: Princeton University Press. 1984

Smith, Nicholas D., trans. "Lesser Hippias", in *Plato: Complete Works*. ed. John M. Cooper and D.S. Hutchinson. 922-36. Indianapolis: Hackett. 1997

Woodruff, Paul, trans. "Greater Hippias", in *Plato: Complete Works*. ed. John M. Cooper and D.S. Hutchinson. 898-921. Indianapolis: Hackett. 1997

The Problem of Theological Fatalism

Matthew Duvalier McCauley
Johns Hopkins University

"Does God know or does He not know that a certain individual will be good or bad? If thou sayest 'He knows', then it necessarily follows that [that] man is compelled to act as God knew beforehand he would act, otherwise God's knowledge would be imperfect…." – Maimonides[1]

INTRODUCTION

Most philosophers and theologians have held that God is omniscient. Many have even argued that if God is omniscient, then he has infallible foreknowledge of the future[2]. God knew the decisions you would make, and every other detail about your life, all from eternity past. "Your eyes saw my unformed body," writes the psalmist; "all the days ordained for me were written in your book before one of them came to be"[3]. If all of our days are written in God's book, then it seems that every human action was, is, and will be fated to occur. No one acts freely. Because of God's infallible foreknowledge, everything everyone does is necessary[4]. This is the thesis of theological fatalism: that divine omniscience and human free will[5] are incompatible.

I will address the problem of theological fatalism by showing why I do not think two of the most prominent contemporary arguments for the thesis are successful.

THEOLOGICAL FATALISM

The Basic Argument

Nelson Pike articulates the problem of theological fatalism in his paper "Divine Omniscience and Voluntary Action."[6] Since his paper has spurred the most discussion in the contemporary literature, I will begin by addressing his version of the argument.

Pike gives the following scenario of a certain 'Jones', who was fated to mow his lawn last Saturday:

> Last Saturday afternoon, Jones mowed his lawn. Assuming that God exists and is (essentially) omniscient…it follows that

---

[1] Gorfinkle, Joseph I., trans. "Semonah Perakhim", in *The Eight Chapters of Maimonides on Ethics*. ed. Joseph I. Gorfinkle. 99-100. New York: AMS Press. 1966

[2] That is, for any event E, God has always known that E would happen. Also, God holds no false beliefs.

[3] Psalm 139:16

[4] 'Necessary' in the metaphysical sense. As in, "The law of excluded middle is necessary." (Not, "Water is necessary for survival.")

[5] I mean "free will" in the libertarian sense.

[6] Pike, Nelson. Divine Omniscience and Voluntary Action. *The Philosophical Review*. 74(1): 27-46. 1965

> eighty years prior to last Saturday afternoon, God knew (and thus believed) that Jones would mow his lawn at that time. But from this it follows, I think, that at the time of action (last Saturday afternoon) Jones was not *able* – that is, it was not *within Jones's power* – to refrain from mowing his lawn. If at the time of action, Jones had been able to refrain from mowing his lawn, then (the most obvious conclusion would seem to be) at the time of action, Jones was able to do something which would have brought it about that God held a false belief eighty years earlier. But God cannot in anything be mistaken…. Thus, last Saturday afternoon, Jones was not able to do something which would have brought it about that God held a false belief eighty years ago.[7]

In other words, God always knew (and thus believed) that Jones would mow his lawn last Saturday. Thus, it was not within Jones' power to refrain from mowing his lawn, since – if he did not mow his lawn – God would have held a false belief about Jones (which is impossible). From this it follows that Jones was not free to refrain from mowing his lawn. He was fated to do it.

We can extend this scenario to include every action that every human performs, and conclude that no human action is voluntary:

1) At time T_2, person P performed action A.
2) If God is omniscient, then at time T_1, God knew and thus believed correctly that at T_2, P would perform A.
3) God is omniscient.
4) Therefore, at T_1, God knew and thus believed correctly that at T_2, P would perform A.
5) Therefore, at T_2, P was not able to perform ¬A.
6) Therefore, the performance of A was involuntary.[8]

Now, I am not so sure that this argument works. It seems that there is a possible world[9] in which God is omniscient, and action A is not fated to happen: namely, that world in which the event described by A does not obtain. If P does not

---

[7] *Ibid*., 32

[8] Harry Frankfurt argued in 1969 that freedom does not imply the ability to do otherwise. That is, simply because P is not free to perform ¬A, it does not follow that the performance of A is involuntary. Since Pike does not acknowledge this argument in his paper, neither will I. At any rate, we can simply replace "involuntary" with "fated to happen."

[9] As used in this sentence, the word "world" does not refer to planet earth or any other part of the physical universe. On the contrary: By "world" I mean "the way the universe is." My definition of "universe" includes the physical, the non-physical, and the *a priori*. By "possible world" I mean "the way the universe could have been."

perform A, it does not follow that God holds a false belief. Instead, all that follows, I think, is that God would have believed something different. If P had performed ¬A at T_2, then God would have believed "P will do ¬A at T_2". God's knowledge that P will perform A is temporally prior to the actual performance of A, but the truth of the proposition "P will perform A at T_2" is logically prior to the fact that God knows it. God's knowledge of that proposition depends on the truth of the proposition, not vice versa.

> Pike seems to anticipate this response. He writes:
> [This response] cannot be accepted. Last Saturday afternoon, Jones was not able to do something that would have brought it about that God believed otherwise than He did eighty years ago…. And if God [believed that Jones would mow his lawn] eighty years prior to Saturday, Jones did not have the power on Saturday to do something that would have made it the case that God did not hold this belief eighty years earlier.[10]

This is all very true, but how does it address my response? Surely, if at T_1 God believed that Jones would mow his lawn at T_2, then Jones can do nothing to change the fact that God believed that Jones would mow his lawn. But this does not show that Jones was fated to mow his lawn. God believes the proposition "Jones will mow his lawn at T_2" *because* the proposition is true. If at T_2, Jones mows his lawn, then – of course! – Jones can do nothing to change the fact that God believed that at T_2 that Jones would mow his lawn. Once an action is performed, there is nothing that one can do to change the fact that the action was performed. But that does not mean that the action in question is fated. Suppose I bake a cake at T_1. Although at T_2 there is nothing I can do to change the fact that I baked a cake at T_1, it hardly follows that my decision to bake a cake at T_1 was involuntary. Similarly, God believes that Jones will mow his lawn at T_2 because it is true that Jones will mow his lawn at that time; and thus Jones can do nothing to change God's belief, since after the time of action, Jones cannot change the fact that he mowed his lawn. Jones' action is not fated. So, again, I do not think that Pike's argument for theological fatalism succeeds.

The Modal Argument

Nonetheless, given advances in contemporary logic, this argument has received a modal logical formulation that circumvents my response to Pike.

[10] *Ibid.*, 32-33

The argument runs as follows. Since God is omniscient, yesterday he infallibly believed that T. Now, if we suppose that if an action occurs in the past, then it is 'now-necessary' that the action in question occurred, then it is now-necessary that yesterday God believed that T. But, it is also necessary that if yesterday God believed that T, then T. So, it follows that it is now-necessary that T. One may then replace "T" with any action, and thus show that every action is fated to occur. Linda Zagzebski outlines this argument in her article, "Foreknowledge and Free Will":[11]

(1) Yesterday God infallibly believed *T.* [Supposition of infallible foreknowledge]
(2) If *E* occurred in the past, it is now-necessary that *E* occurred then. [Principle of the Necessity of the Past]
(3) It is now-necessary that yesterday God believed *T.* [1, 2]
(4) Necessarily, if yesterday God believed *T*, then *T.* [Definition of "infallibility"]
(5) If *p* is now-necessary, and necessarily $(p \rightarrow q)$, then *q* is now-necessary. [Transfer of Necessity Principle]
(6) So it is now-necessary that *T.* [3,4,5]
(7) If it is now-necessary that *T*, then [Jones cannot do otherwise than mow his lawn Saturday afternoon]. [Definition of "necessary"]
(8) Therefore, [Jones cannot do otherwise than mow his lawn Saturday afternoon]. [6, 7]
(9) If you cannot do otherwise when you do an act, you do not act freely. [Principle of Alternate Possibilities]. Therefore, when [Jones mows his lawn Saturday afternoon, he] will not do it freely.[12]

We can formalize Zagzebski's argument as follows:

1) Yesterday, $B_{G(infallibly)}T$ [Supposition of infallible foreknowledge]
2) $Ep \supset \Box E$ [Principle of the Necessity of the Past]
3) $\Box(yesterday)B_GT$ [1, 2]
4) $\Box(B_GT \supset T)$ [Definition of "infallibility"]
5) $(\Box(B_GT \supset T) \wedge \Box(yesterday)B_GT) \supset \Box$ T [Transfer of Necessity Principle]

[11] Zagzebski, Linda. Foreknowledge and Free Will. *The Stanford Encyclopedia of Philosophy*. ed. Edward N. Zalta. Fall 2011
[12] *Ibid.,* 1

6) □T [3,4,5]
7) □T⊃(Jones cannot do otherwise than mow his lawn) [Definition of "necessary"]
8) Jones cannot do otherwise than mow his lawn [6,7]
9) Therefore, when Jones mows his lawn, he does not do it freely

If all the premises in this argument are true, then it follows that Jones is not free to mow his lawn. Again, this observation can be extended to any action performed by any person. Notice that this modal version of theological fatalism circumvents my response to Pike's argument. Since it is now-necessary that God believed at T_1 that Jones would mow his lawn at T_2, then the proposition "Jones will mow his lawn at T_2" is necessary. In order to respond to this argument, we will have to reject one of the premises. To this we now turn.

Against the Principle of the Necessity of the Past

My aim in this section is to refute the Principle of the Necessity of the Past (PNP), found in premise (2) of the above argument. I will argue 1) that PNP is conceptually problematic, and 2) that PNP is logically fallacious as well. To defend this second contention, I will need to craft a modal language for metaphysics, that is, a modal language that will enable us to address metaphysical concerns, such as the one before us.

Now, it seems to me that PNP is conceptually problematic. When you think about the principle, it does not make much sense. What exactly does it mean to say that it is 'now-necessary' that some event in the past happened? That the event in question obtains in every possible world? That clearly cannot be true, since for any contingent event E, it is possible that that event did not occur. So 'now-necessary' must mean something else. Does it mean that a now-necessary event cannot be changed? Perhaps. Events in the past – as with *a priori* truths – are unchangeable. For example, we can no more change the fact that *Halo 2* won Console Game of the Year in 2005, than we can change the Law of Excluded Middle. One might conclude, therefore, that since necessary truths are unchangeable, unchangeable things are necessary. So, the past is necessary because it is unchangeable.

But it is unclear that it follows from the fact that necessary truths are unchangeable, that the (unchangeable) past is necessary. There are indeed very many things that we cannot change that are nonetheless not necessary. The moon, for example, encircles the earth with a mean orbital velocity of 1,023

meters per second, and there is nothing that we can do to change that fact. And yet, we would hardly say that the orbit of the moon around the earth is necessary. My point is that it is unreasonable to conclude the necessity of the past simply by observing the fact that the past cannot be changed.

But beyond being conceptually mysterious, PNP is logically fallacious. To demonstrate this, I will need to construct a modal language for metaphysics. I think it should look something like the following[13].

Let us suppose that the metaphysical modal language Lm extends the basic sentential modal language with the *contingency* operator (directly below, sixth column):

$$P \mid \bot \mid \neg\phi \mid (\phi \wedge \phi) \mid \Box\phi \mid \Diamond_p\phi \mid \Diamond_c\phi$$

Read $\Box\phi$ as 'It is necessary that ϕ' and $\Diamond p\phi$ as 'It is possible that ϕ'. Read $\Diamond_c\phi$ as 'It is *contingent* that ϕ'. The semantics for L_m is the standard one based on Kripke models where *wRv* just in case *v* is a *metaphysically possible* world relative to *w*.

Let us assume that it is a metaphysical principle that the modality of a thing or a proposition cannot change. This is to say that if a thing or a proposition is merely possible (that is, possible, but not necessary), then its modality cannot change from merely possible to necessary; it is necessarily merely possible. If a thing or a proposition is necessary, then it cannot change from necessary to merely possible; it is necessarily necessary. This stipulation will require us to work with a system at least as strong as the modal system K45 (defined below). So, for present purposes, we will construct the *Metaphysical Logic (ML)* as the logic of K45:

(PL) All (substitutions of) tautologies are axioms
(MP) From ϕ and $\phi \supset \psi$ infer ψ
(Nec_m) From ϕ infer $\Box\phi$
(K_m) For any ϕ, ψ, $\Box(\phi \supset \psi) \supset (\Box\phi \supset \Box\psi)$
(4_m) For any ϕ, $\Box\phi \supset\Box\Box\phi$ is an axiom.
(5_m) For any ϕ, $\Diamond_p\phi\supset\Box\Diamond_p\phi$ and $\Diamond_c\phi\supset\Box\Diamond_c\phi$ are axioms

A model M = <W, R, V> consists of a nonempty set W of worlds, a valuation function $V : At_L \times W \rightarrow \{T, F\}$, and a binary accessibility relation $R \subseteq W \times W$ between worlds. The following recursive clauses lift V to the complete interpretation function $[[\ ]]_M : S_L \times W \rightarrow \{T, F\}$ for L_m:

[13] I will henceforth use logical concepts and terminology; thus it is assumed that the reader has a basic understanding of modal logic. My aim is not to confuse, but to argue definitively against PNP. This will require a bit of technicality.

$[[P]]^{M}_{w}$ = T iff V(P,w) = T
$[[\perp]]^{M}_{w}$ = T iff 0=1
$[[\neg\phi]]^{M}_{w}$ = T iff $[[\phi]]^{M}_{w}$ = F
$[[(\phi \wedge \psi)]]^{M}_{w}$ = T iff $[[\phi]]^{M}_{w} = [[\psi]]^{M}_{w}$ = T
$[[\Box\phi]]^{M}_{w}$ = T iff $\forall v \in \{v : wRv\}([[\phi]]^{M}_{V}$ = T)
$[[\Diamond_{p}\phi]]^{M}_{w}$ = T iff $\exists v \in \{v : wRv\}([[\phi]]^{M}_{V}$ = T)
$[[\Diamond_{c}\phi]]^{M}_{w}$ = T iff $\exists v \in \{v : wRv\}([[\phi]]^{M}_{V}$ = T) and $\neg(\forall v \in \{v : wRv\}$ $([[\phi]]^{M}_{V}$ = T))

In other words, $\Diamond_{c}\phi$ – that is, 'It is contingent that ϕ' – is true in M at a world w just in case 1) there is at least one world accessible from w where ϕ is true, and 2) ϕ is not true at all worlds accessible from w. This captures the sense of 'merely possible', distinct from $\Diamond_{p}\phi$.

Now, given L_{m} we can address the modal version of the argument for theological fatalism. Let us take another look at premise (2), $Ep \supset \Box E$ – that is, if E occurred in the past, then E is now necessary. If before E occurred, E was *merely* possible, then E has changed its modality from contingent to necessary, that is:

$$\Diamond_{c}\phi \supset \Diamond_{p}\Box\phi$$

must be valid. Using our metaphysical modal language, we can show that it is in fact not valid, and, therefore, that PNP is invalid as well (since in order for PNP to get off the ground, it must be possible for the modality of a proposition to change from merely possible to necessary). Let A be any action. Then:

1) $\Diamond_{c}A \supset \neg\Box A$ [follows from semantics of $\Diamond_{c}A$]
2) $\Diamond_{c}A$ [given]
3) $\neg\Box A$ [1,2,mp]
4) $\Diamond_{c}A \supset \Diamond_{p}\Box A$ [follows from principle of necessity of the past]
5) $\Diamond_{p}\Box A$ [2,4,mp]
6) $\Diamond_{p}\neg A \supset \Box\Diamond_{p}\neg A$ [5 axiom, $\neg A$ instantiation]
7) $\neg\Box\Diamond_{p}\neg A \supset \neg\Diamond_{p}\neg A$ [contrapositive,6]
8) $\Diamond_{p}\Box A \supset \Box A$ [duality,7]
9) $\Box A$ [5,8,mp] ***(contradicts 3)***

Since the contradiction of (3) and (9) is a direct consequence of $\Diamond_{c}\phi \supset \Diamond_{p}\Box\phi$, our metaphysical language rules this out. And since $\Diamond_{c}\phi \supset \Diamond_{p}\Box\phi$ is a direct consequence of PNP, PNP is ruled out as well. Premise (2) of the modal argument of theological fatalism is invalid.

Once we reject premise (2), we can see how the modal argument for theological fatalism founders. Recall, if it is necessary that at time T_1, God believed that at T_2, person P would perform action A, then A, as performed by P, is necessary. However, since we have refuted PNP, we no longer have a reason to think that it is necessary that God believed P would perform A. As illustrated in my response to Pike's argument, the performance of A – and God's belief that A will be performed – is a contingent fact. ¬A could have been performed, and if the proposition 'P will perform ¬A at time T_2' were true, then God would have believed it. Thus, even the modal version of the argument from theological fatalism is not successful.

CONCLUSION

If my paper is successful, then the above two arguments for theological fatalism are invalid. But this is hardly the entire job. For one, there might be modal versions of the argument that do not rely on PNP. If there are, then they would avoid my responses. Second, merely refuting the arguments for theological fatalism *does not* establish that the thesis itself (i.e., that divine omniscience and human free will are incompatible) is false. To do that, we would need to construct an account of God's foreknowledge that entails that humans are free.

Moreover, there are reasons to think that my aim in this paper is not successful. For one, I have not crafted a formal account of God's omniscience. Perhaps a more careful articulation of omniscience would work against my arguments. Second, one might think that my metaphysical language is somewhat *ad hoc*, and contrived simply for the purpose of responding to theological fatalism (Why accept the metaphysical principle that the modality of a thing or a proposition cannot change?)

I do not have responses to these concerns at the moment, but they are certainly avenues for further research.

BIBLIOGRAPHY

Gorfinkle, Joseph I., trans. "Semonah Perakhim", in *The Eight Chapters of Maimonides on Ethics*. ed. Joseph I. Gorfinkle. 99-100. New York: AMS Press. 1966

Pike, Nelson. Divine Omniscience and Voluntary Action. *The Philosophical Review*. 74(1): 27-46. 1965

Zagzebski, Linda. Foreknowledge and Free Will. *The Stanford Encyclopedia of Philosophy*. ed. Edward N. Zalta. Fall 2011

Manipulation, Argument, & Experiment:

Putting Folk Intuitions into Context

Matthew Paskell
The University of Arizona

1. MANIPULATION AND THE FOUR-CASE ARGUMENT[1]

Manipulation arguments seek to describe an agent who is manipulated either to act in a particular way or to have certain tendencies that play a significant causal role in actions. The agent, though manipulated, is also meant to satisfy what McKenna has called a Compatibilist-friendly Agential Structure (CAS) (McKenna 2008). This structure includes the freedom-relevant conditions that a compatibilist would hold to be minimally sufficient for the moral responsibility of an action that is the direct product of CAS. Since there is no unique set of conditions that all compatibilists agree are minimally sufficient, both the content of CAS and the relevant psychological features of an agent in a manipulation argument vary. The goal of a manipulation argument then is to devise a scenario in which a manipulated agent acts from CAS and *cannot* be said to be morally responsible for that action. If the manipulation argument is both properly constructed and gives rise to the belief that the agent is not morally responsible, then CAS is said to be insufficient. Since CAS is intended to justify holding agents morally responsible in a determined world, a successful manipulation argument would give reason to doubt that moral responsibility is compatible with determinism.

The most prominent manipulation argument is perhaps Derk Pereboom's "Four-Case Argument," outlined in his *Living Without Free Will*. Pereboom not only develops a manipulation argument but also combines this with a generalization strategy—by starting with a case of covert manipulation and moving through a set of increasingly more "natural" cases, he attempts to show that there is no relevant difference between a case of manipulation and his final case, an ordinary determined agent. What is also unique to Pereboom's argument is that the CAS his agents are meant to satisfy incorporates four different compatibilist accounts—that of Ayer (1954), Frankfurt (1971), Fischer and Ravizza (1998), and Wallace (1994). This means that, if his argument is in fact successful, then all of these compatibilist accounts can be shown insufficient simultaneously. Pereboom begins by describing a case in which a manipulated agent, Professor Plum, kills Ms. White:

> **Case 1**. Professor Plum was created by neuroscientists, who can manipulate him directly through the use of radio-like technology, but he is as much like an ordinary human being as is possible, given his history. Suppose these neuroscientists "locally" manipulate him to undertake the process of reasoning

[1] I am incredibly grateful to Michael McKenna, both for reviewing multiple drafts of this work and for many beneficial discussions about free will and moral responsibility.

> by which his desires are brought about and modified - directly producing his every state from moment to moment. The neuroscientists manipulate him by, among other things, pushing a series of buttons just before he begins to reason about his situation, thereby causing his reasoning process to be rationally egoistic. Plum is not constrained to act in the sense that he does not act because of an irresistible desire—the neuroscientists do not provide him with an irresistible desire - and he does not think and act contrary to character since he is often manipulated to be rationally egoistic. His effective first-order desire to kill Ms. White conforms to his second-order desires. Plum's reasoning process exemplifies the various components of moderate reasons-responsiveness. He is receptive to the relevant patterns of reasons, and his reasoning process would have resulted in different choices in some situations in which the egoistic reasons were otherwise. At the same time, he is not exclusively rationally egoistic since he will typically regulate his behavior by moral reasons when the egoistic reasons are relatively weak - weaker than they are in the current situation. (Pereboom 2001, pg.112-3)

In this case Professor Plum is said to have satisfied the relevant CAS, what Pereboom calls the "four causal integrationist conditions." His desire to kill White is consistent with his character, he has a second-order volition that conforms to this desire, and he rationally considers relevant reasons and is moderately reason-responsive. Pereboom also suggests that "intuitively, [Plum] is not morally responsible because he is determined by the neuroscientists' activities, which are beyond his control" (2001, pg. 113). While this alone may be a plausible candidate for a successful manipulation argument, Pereboom describes another, slightly different case both to answer a potential objection to Case 1 and to advance his generalization strategy:

> **Case 2**. Plum is like an ordinary human being, except that he was created by neuroscientists, who, although they cannot control him directly, have programmed him to weigh reasons for action so that he is often but not exclusively rationally egoistic, with the result that in the circumstances in which he now finds himself, he is causally determined to undertake the moderately reasons-responsive process and to possess the set of first- and second-order desires that results in his killing Ms.

> White. He has the general ability to regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and accordingly he is causally determined to kill for these reasons. Nevertheless, he does not act because of an irresistible desire. (Pereboom 2001, pg.113-4)

This case is similar to Case 1, though rather than being directly manipulated by the neuroscientists, Plum is programmed at an earlier time to have the rationally egoistic tendencies he does. This answers a possible objection to Case 1—that somehow the responsibility-undermining condition has to do with Plum being "locally manipulated." Using Case 2 Pereboom argues that, whether or not there is a "time lag" between the neuroscientists' commands and Plum's actions, it seems that Plum is not morally responsible. The timing of the manipulation, then, does not seem to be a relevant factor. From here a third case is introduced, one in which cultural conditioning takes the place of the neuroscientists:

> **Case 3**. Plum is an ordinary human being, except that he was determined by the rigorous training practices of his home and community so that he is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1 and 2). His training took place at too early an age for him to have had the ability to prevent or alter the practices that determined his character. In his current circumstances, Plum is thereby caused to undertake the moderately reasons-responsive process and to possess the first- and second-order desires that result in his killing White. He has the general ability to grasp, apply, and regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and hence the rigorous training practices of his upbringing deterministically result in his act of murder. Nevertheless, he does not act because of an irresistible desire. (Pereboom, 2001 pg. 114)

Here we have a scenario where CAS is again met, though the likelihood that this case, taken by itself, will elicit the intuition that the agent is not morally responsible is less than in the previous two cases. After all, we can find scenarios like this in the real world and we don't always, or perhaps even often, judge them to be responsibility-undermining. However we might try to account for our attitudes toward this case, though, there will be a challenge—there seem to be no responsibility-relevant differences between this case and the previous two. If we accept that Plum is not morally responsible in those cases then

we must make the same judgment about Plum in Case 3. To complete his generalization, Pereboom outlines a final case, Plum as an ordinary person in a deterministic world:

> **Case 4.** Physicalist determinism is true, and Plum is an ordinary human being generated and raised under normal circumstances, who is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1-3). Plum's killing of White comes about as a result of his undertaking the moderately reasons-responsive process of deliberation, he exhibits the specified organization of first-order and second-order desires, and he does not act because of an irresistible desire. He has the general ability to grasp, apply, and regulate his behavior by moral reasons, but in these circumstances the egoistic reasons are very powerful, and together with background circumstances they deterministically result in his act of murder. (Pereboom 2001, pg. 115)

When considering Case 4 we must again ask whether there are any responsibility-relevant differences between this case and the previous ones. If we accept that the first three cases successfully captured CAS then the answer is simple. Compatibilists will accept that an ordinary determined agent also satisfies these conditions because they accept both that determinism may be true and that one or more of these conditions allow for moral responsibility in a determined world. If we also accept that one of the Plums in the first three cases is not morally responsible then we should hold the same attitude toward the ordinary determined Plum.

Pereboom's conclusion is that the first three Plums are indeed nonresponsible, that CAS is satisfied in each of these cases, and thus the conditions in CAS are insufficient for holding agents moral responsibility. What is missing, according to Pereboom, is what he calls the Causal History Principle—"An action is free in the sense required for moral responsibility only if the decision to perform it is not an alien-deterministic event, nor a truly random event, nor a partially random event" (Pereboom 2001, pg. 54). Since this condition is not satisfied by ordinary determined agents, moral responsibility is incompatible with determinism.

2. OBJECTING TO THE FOUR-CASE ARGUMENT

There are two ways in which a compatibilist might object to Pereboom's four-case argument and try to avoid the conclusion that an ordinary determined

agent is not morally responsible. The first is a soft-line response, which would hold that the manipulation cases, as described, do not successfully capture CAS. If this is true then any belief that Plum is not morally responsible in cases 1 or 2 can be explained by an appeal to whatever conditions may be missing. It could be argued, for example, that Plum's manipulation prevents him from having the type of identity-forming history that a morally responsible agent should have (Pereboom 2001, pg.120-1). There may also be concern that Plum is being manipulated by other agents and so the fact that Plum is "a puppet of another person" is what undermines his responsibility (Pereboom 2001, pg. 115).

The problem with this type of response is that the scenario can always be adjusted so that any additional conditions a compatibilist finds necessary for responsibility are satisfied by the agent. Pereboom, in addition to the initial four compatibilist conditions, can simply tack on another feature to Plum's psychology. The soft-line approach is always likely to fail since, as McKenna points out, there seems to be "no way to foreclose the metaphysical possibility that the causes figuring in the creation of a determined morally responsible agent could not be artificially fabricated" (McKenna 2008, pg. 144). Even if the compatibilist argues that a manipulation case has yet to be constructed properly, it would be quite a leap to suggest that a manipulated agent who satisfies any and all compatibilist conditions for responsibility is *inconceivable*. After all, what factor could we point to in a case of a manipulated agent that could not also arise in a determined world? The primary difference between a relevantly described manipulated agent and an ordinary determined agent seems to lie in the source of determination rather than in the type of determination, and the source of determination could hardly be relevant to responsibility. A soft-line reply, then, would not be a promising approach to countering Pereboom's four-case argument. The compatibilist will need to take a different route.

While the soft-line reply denies that the manipulated agent satisfies CAS, the hard-line reply accepts that CAS is fulfilled and embraces that fact. This can take one of two forms, each aimed at denying Pereboom his jumping-off point—the nonresponsibility of Plum in Case 1. First, the ambitious hard-liner may give a positive argument in an attempt to establish that Plum, or an agent in comparable circumstances, is in fact a morally responsible agent. One way that this approach is ambitious is that any argument for Plum's responsibility in Case 1 will have to overcome the initial intuition that extreme, covert manipulation threatens moral responsibility. This would not only require casting doubt on Plum's nonresponsibility, but making a case strong enough that inclines us toward Plum's being responsible. The other way that this approach is ambitious is that giving a positive argument for Plum's responsibility is more than what is necessary for defeating Pereboom's argument. In order to successfully

counter the four-case argument all that the compatibilist needs to do is show that Plum's nonresponsibility in Case 1 cannot be taken for granted. The other approach a hard-liner can take is just this—showing that, at the *very least*, Plum's nonresponsibility is not a given. This is the position that McKenna develops and, in outlining his argument against granting Plum's nonresponsibility, he uses Pereboom's generalization strategy against him.

McKenna begins his hard-line reply by suggesting that, rather than offering a soft-line reply to a manipulation argument that falls just short of capturing CAS, compatibilists should take it upon themselves to improve the argument. With respect to Pereboom's cases, then, McKenna adds that each Plum has a proper history and is "'morally articulate' in a broadly Strawsonian fashion" (McKenna 2008, pg. 152).[2] The next step is to draw attention to the fact that a manipulated agent who satisfies CAS, "lives up to a rich sort of agency and genuinely satisfies certain moral properties" (McKenna 2008, pg. 144). Drawing attention to the ways in which a properly described manipulated agent is similar to an ordinary agent can lessen the intuitive force behind Plum's nonresponsibility in the Case 1. It can also help clarify what initial attitude it is rational to hold toward Plum in Case 1 and McKenna illustrates both of these ideas by considering Pereboom's generalization in reverse-order.

McKenna begins with Pereboom's Case 4: Plum as an ordinary determined agent. Since the responsibility or nonresponsibility of an ordinary determined agent is exactly what is at issue, it cannot be claimed outright that Plum in Case 4 is not morally responsible. This, of course, would be question begging and so the appropriate initial attitude will be some form of agnosticism. Since all relevant features of Plum are meant to carry over in all cases, the appropriate initial attitude that is informed by those features should also be preserved. Thus if we run the generalization backwards, from Case 4 through to Case 1, then the attitude toward Plum in Case 1 should also be some form of agnosticism.[3] This helps clarify the initial attitude we ought to hold toward Plum in Case 1 but it also does more than this—the order in which the cases are presented can determine which features of those cases are emphasized. Examining the cases in Pereboom's order highlights the ways in which determinism can be analogous to certain forms of manipulation. This is something we ought to take into account, no doubt, but examining the cases in McKenna's order highlights the ways in which a manipulated agent can have

[2] What it means to be "morally articulate" can be outlined in more detail, as McKenna does (2008, pg. 151). The general feature will be that Plum views himself, and others view him, as a proper subject of moral reactive attitudes.

[3] To make the generalization even clearer, McKenna adds two additional cases, one in which God determines the complete state of the world and one in which a deity, Diana, meddles with Plum as a zygote. In both cases the source of "manipulation" is changed but the other events, including Plum's psychology, unfold just as they do in Pereboom's four-cases. (McKenna, 2008, pg. 152-3)

all the moral features of a normal agent. This is also something we should consider in our judgments about each Plum and it makes it more difficult to declare outright that the manipulated Plums are not responsible. Examining the cases in this order may not give rise to the belief that the manipulated Plum's *are in fact* responsible, but this does not need to be the compatibilist's goal. Simply casting doubt on Plum's nonresponsibility is enough. This is because the compatibilist "needs only to show that the incompatibilists who advance the Manipulation Argument are not clearly right about the cases they feature to establish a key premise of their argument" (McKenna 2008, pg. 155). A key premise in Pereboom's four-case argument is that Plum in Case 1 is not morally responsible, and so a hard-line reply like McKenna's is successful so long as it shows that this is not clearly the case.

At this point in the debate, McKenna concludes that the result is a dialectical stalemate—which is a victory for the compatibilist—and Pereboom, in addressing McKenna's argument, holds that the intuitions would still fall in his favor (McKenna 2008, pg. 154). While there is more that will be said about each of these positions, as both McKenna and Pereboom have developed them further, there is another way that some have attempted to challenge both manipulation arguments and the four-case argument in particular. Before returning to the McKenna-Pereboom debate it will be helpful to address the claims made by these empirical approaches.

3. FOLK INTUITIONS AND MANIPULATION

Experimental philosophy is a growing field and folk intuitions about free will have been a central area of research. With manipulation arguments being a major focus of the current debate between compatibilists and incompatibilists, it is no surprise that some philosophers have attempted to examine folk intuitions about manipulated agents. In Chandra Sekhar Sripada's article "*What Makes a Manipulated Agent Unfree?*", he attempts to discern which aspects of manipulation cases drive our judgments about responsibility. While Sripada's experiment does not test Pereboom's arguments directly, the conclusions he draws are meant to have implications for all manipulation arguments. Adam Feltz, on the other hand, designed an experiment intended to assess intuitions about Pereboom's four cases, both individually and in sequence. Feltz' goal is to determine the success of the four-case argument by using folk intuitions to address its key premises. Both Sripada and Feltz suggest that folk intuitions support the compatibilist position, and that their results can support both a soft-line and hard-line reply. Taking these experiments one at a time, however, we will see that neither is entitled to this conclusion.

With his experiment, Sripada attempts to answer the question, "What features of manipulation cases are our intuitions sensitive to?" (Sripada 2012, pg. 565). Since incompatibilists argue that a lack of ultimate control drives judgments of non-responsibility and compatibilists argue that it is features of the agent's psychology, studying the responses to cases of manipulation may give insight into which of these positions is correct. Sripada's experiment was designed to test the plausibility of what he calls the "Compatibilist Position," which he divides into three claims:

> 1. Intuitions in manipulation cases are sensitive to whether or not there is damage to the manipulated agent's psychological capacities (due to factors such as corrupted information and deep self discordance).
> 2. (Soft-line) To the extent that the manipulated agent is seen as exhibiting damaged psychological capacities, the agent is intuitively judged to be unfree.
> 3. (Hard-line) To the extent that the manipulated agent is not seen as exhibiting damaged psychological capacities, the agent is intuitively judged to be free. (Sripada 2012, pg. 573)

If the Compatibilist Position is correct then it is psychological factors that drive judgments of non-responsibility and we would expect intuitions about manipulation cases to reflect these three claims.

Sripada's participants were asked to read a vignette about a man named Bill and a neuroscientist, Dr. Z:

> One day, Bill sees a woman named Mrs. White as she is jogging in the park. Bill hates this woman, and deliberates about what to do. After weighing his options, Bill decides he should kill her. Bill's mind is not clouded by rage or other extreme emotions. Rather, Bill thinks clearly and carefully about his own desires and values, and only then makes a decision. After he kills Mrs. White, Bill reflects on his action. He wholeheartedly endorses what he has done.
>
> But there is more you need to know about Bill, and how he came to be the person that he is now:
>
> There is a man named Dr. Z who is a scientific genius and

> who is an expert at indoctrination. Dr. Z hates Mrs. White and formed the following plan. Dr. Z would take an infant from an orphanage and raise the child himself. He would teach and reward just the right behaviors in the child so the child would hate Mrs. White and want her dead. He would script all the major events in the child's life to nurture and cultivate in the child the goal of doing whatever it would take to kill Mrs. White. Dr. Z tried this plan previously on five other children, and each time the child grew up to kill Dr. Z's intended targets. (Sripada 2012, pg. 573-4)

Half of the participants were put in the "Manipulation" condition and told that Dr. Z adopted Bill and that his indoctrination worked. The other half were in the "No Manipulation" condition and were told that Bill was adopted by someone else and thus was not influenced by Dr. Z. In both cases Bill grew up to hate and kill Ms. White. All participants were then asked to rate their agreement with a series of statements on a scale from 1 to 7:[4]

1. Bill killed Mrs. White of his own free will.
2. Bill was in control of whether or not he killed Mrs. White.
3. Bill is morally responsible for killing Mrs. White.
4. Bill killed Mrs. White based on false information about her, and he was deprived of any opportunity to learn the truth.
5. Bill was never taught about why certain actions are right and wrong, so he does not truly know that killing Mrs. White is wrong.
6. Bill killed Mrs. White because his upbringing kept him ignorant of alternative, non-violent, ways of acting.
7. Bill's killing of Mrs. White does not reflect the kind of person who he truly is deep down inside.
8. The real Bill did not truly want to kill Mrs. White—Bill killed only because Dr. Z wanted him to.
9. Bill is constrained by Dr. Z to act in a way that differs from how he himself, deep down, wants to act. (Sripada 2012, pg. 575, Table 1)

Sripada divides these statements into three groups: 1-3 assess Bill's Free Will rating, 4-6 assess whether Bill is seen as having Corrupted Information, and 7-9

[4] A rating of 1indicated that the participant "Strongly Agreed" while 7 meant "Strongly Disagree." In his analysis Sripada used scores of 1, 2, or 3 as expressing agreement (Sripada, 2012, pg. 576, fn 11).

measure whether Bill has Deep Self Discordance. The Compatibilist Position predicts that those who express disagreement with the Free Will statements (1-3) will agree with the statements suggesting damaged psychological capacities (4-9), and vice versa. The analysis of the results of this study agreed with the Compatibilist Position's prediction—those who saw Bill as unfree and not responsible tended to think that he was psychologically damaged, and those who rated Bill free and responsible did not think that he was psychologically damaged. As Sripada notes, though, there was an even stronger relationship:

> "To the extent that people thought that Bill suffered from deep self discordance or corrupted information, or both, they tended to think Bill *lacked* free will in killing Mrs. White. While to the extent that people thought Bill did *not* suffer these afflictions, they tended to think that Bill *possessed* free will in killing Mrs. White." (Sripada 2012, pg. 584)

So not only were free will and responsibility ratings inversely related to ratings of psychological damage, the higher a participant rated one category the lower they rated the other. This is said to support a soft-line reply to manipulation arguments since judgments that Bill is unfree and not responsible are explained by appealing to psychological features. That is, the corrupted information and deep-self discordance prevent the agent from satisfying CAS and *this* is what drives the nonresponsibility intuitions. These results are also meant to support a hard-line reply to manipulation arguments. This is because, when Bill was not seen as psychologically damaged, and hence satisfied CAS, he was judged free and responsible. Thus Sripada's results demonstrate that folk intuitions match his Compatibilist's Prediction, but does this force us to conclude that people have compatibilist intuitions about manipulation cases?

One reason to doubt this conclusion is that the questions Sripada asked participants might not reflect compatibilist judgments so much as they *coax* compatibilist judgments. We should recall that questions 1-3 are used to assess judgments of Bill's freedom and responsibility while questions 4-9 are meant to be potential explanatory factors for these judgments. The problem is that there is nothing in the vignette that would allow a participant to justify an answer to any of these corrupted information or deep self discordance questions. The only way that a participant could determine that Bill's "upbringing kept him ignorant of alternative, non-violent, ways of acting," or that he is "constrained by Dr. Z to act in a way that differs from how he himself, deep down, wants to act," is if they read into the vignette what was simply not there.

Sripada, giving what could be taken as a defense of these questions, suggests

both that manipulation presumably *must* involve corrupted information and that a *tabula rasa*, or *blank slate* view of the human mind is "surely unlikely to be *thought* to be true by most people" (Sripada 2012, pg. 570). The participants, then, should be justified in inferring that Bill received corrupted information and that he has a "deep self" which is distinct from what Dr. Z manipulated him to be. The problem with this defense is that, in manipulation arguments, the manipulator is meant to be analogous to deterministic processes. In any plausible manipulation argument the agent will not receive any more corrupted information than an ordinary person might. Likewise, there will be no self which is distinct from the manipulated self any more than there would be a self distinct from a person's "determined self." As we have seen, there is no reason to think that the features present in a determined agent could not also be brought about through manipulation, and allowing participants to *presume* that CAS is not satisfied is to deny this without argument. If the claim is simply that it is too difficult for most people to imagine a manipulated agent who is otherwise "normal," or who satisfies CAS, then this is not an indictment against manipulation arguments, but rather a reason for thinking that folk intuitions might be limited in their application.

Another problem with Sripada's conclusion arises from the fact that he chose to test only the Compatibilist Position. We can note at the outset that, just because the results of the study matched the Compatibilist Position, this does not guarantee that it is only or even primarily psychological factors that drive intuitions about manipulated agents. We should not, however, think that this is what is being argued. Rather, Sripada suggests that corrupted information and deep self discordance "*fully* explained variation in people's free will judgments" and that intuitions "do not track the features that incompatibilists say they track (i.e., the agent's lack of ultimate control over his or her actions)" (Sripada 2012, pg. 582-583). We must ask, though, how it is that Sripada establishes this conclusion. Since only compatibilist explanations for people's free will judgments were examined we cannot know whether an "Incompatibilist Position" would have tracked these judgments equally well. After all, incompatibilist hold that ultimate control is a necessary condition for freedom and responsibility and it may be that psychological damage, or manipulation itself, is seen as undermining this type of control. To support this we can point out that, to the extent that an Incompatibilist Position *was* tested in this study, the results might also be said to support *this* position. In question 2 participants were asked whether "Bill was in control of whether or not he killed Mrs. White." The results were the same as those probing compatibilist explanations—to the extent that he was seen lacking control he was judged unfree and not responsible, and to the extent that he was seen as having control he was judged free and responsible

(Sripada 2012, pg. 589, Appendix Table A1). Of course, we cannot know what type of control the participants thought was undermined, but this is precisely the point— having only tested the predictions of the Compatibilist Position, and having no comparable Incompatibilist Position to contrast with, we cannot come to a conclusion based on these results.[5]

While Sripada's study takes aim at manipulation arguments generally, Feltz attempts to use folk intuitions to address Pereboom's argument directly. In his experiment participants were asked to read the same introductory paragraph as in Sripada's study, describing the man named Bill who kills a jogger, Mrs. White. After reading this paragraph participants were given four descriptions of Bill intended to mimic Pereboom's four cases. The first was what Feltz dubbed "Intentional Direct Manipulation:"

> Bill is essentially a normal man, but he was created by neuroscientists who directly manipulate all of his decisions. The neuroscientists manipulate Bill to make decisions that almost always benefit him. The neuroscientists implant in Bill a desire to kill Mrs. White. He is able to regulate his behavior by moral reasoning and act differently in different situations with different reasons, but in the present circumstances, the desire to kill Mrs. White is stronger than any competing desire. As a result of the neuroscientists' implanting in Bill the desire to kill Mrs. White, Bill decides to kill Mrs. White and does it. Reflecting on the action afterward, Bill identifies with the desire to kill Mrs. White and the resulting action. (Feltz 2012, pg. 56)

In the second case, "Intentional Indirect Manipulation," the neuroscientists do not manipulate Bill locally but have instead programmed his genes. In the third, "Culture" case, Bill was unavoidably trained by his community, and in the final, "Determinism" case, Bill is "completely caused by his genes and his cultural environment" (Feltz 2012, pg. 56). In each of these cases, as in Pereboom's, the agent has the same desires and psychological states. Participants were randomly assigned to either read only the Determinism case or to read all four cases separately and in sequential order. They were instructed to answer six questions on a scale from 1-7, 1 meaning "strongly disagree" and 7 meaning "strongly agree:"

---

[5] Sripada does state that he has studied folk intuitions on ultimate control, and that "people are quite willing to attribute free will and moral responsibility to an agent whose fundamental values and evaluative attitudes are clearly and obviously determined by factors that are completely out of her control." If this is the case then the combination of these studies *may* give more weight to the Compatibilist Position, but since the ultimate control studies are still "in preparation" we cannot declare a compatibilist picture of manipulated agents "intuitive."

1. Bill kills Mrs. White of his own free will.
2. Bill's killing of Mrs. White is "up to him."
3. Bill is morally responsible for killing Mrs. White.
4. Bill is blameworthy for killing Mrs. White.
5. Bill deserves to be punished for killing Mrs. White.
6. Bill should be prevented from killing Mrs. White.

Feltz also conducted a second experiment that was setup in the same way, though rather than neuroscientists manipulating Bill, his psychological features were a result of a brain tumor that either manipulated him directly (Case 1, "Non-Intentional Direct Manipulation") or programmed his decisions (Case 2, "Non-Intentional Indirect Manipulation"). This was to test the idea proposed by Alfred Mele (2006), that intuitions might change based on the intentionality of the manipulator (Feltz 2012, pg. 57).

The results of these experiments indicated that it was only in the case of Intentional Direct Manipulation, Bill in Case 1 of Experiment 1, that Bill was rated unfree and not responsible (Feltz 2012, pg. 59). The study also showed that the closer to "normal" a case was the more people tended to rate the agent free and responsible (Feltz 2012, pg. 59, Fig. 1). Feltz suggests that this can be used in support of either a soft-line reply to Pereboom's four-case argument or a hard-line reply. A soft-line reply might be supported since participants seem to find freedom- and responsibility-relevant differences between manipulated agents and determined agents. It could be argued, then, that there is something undermining moral responsibility in the former that is not present in the latter. Feltz entertains a few explanations for this—the temporal proximity of the manipulator, the fact that Bill is a puppet of another person, and the different levels of psychological damage attributed under different circumstances (Feltz 2012, pg. 60). The results can also support a hard-line reply, according to Feltz, because Bill was generally rated free and responsible in the Non-Intentional Direct Manipulation case. Since Bill was rated not free or responsible in the Intentional Direct Manipulation case, and the difference between these cases is intentionality, it could be argued that changes in intuitions arise from the intentionality rather than the mere presence of manipulation. Presumably, the removal of intentionality results in CAS being satisfied, and it is only then that participants are willing to rate Bill free and responsible. Feltz also follows McKenna in suggesting that our intuitions about real-world cases are more reliable than our intuitions about bizarre cases (McKenna 2008). This would mean that intuitions about the non-intentional brain tumor case are more trustworthy than intuitions about meddling neuroscientists.

Although Feltz' study may give us valuable insight into how the folk

judge manipulated agents, we should be cautious in trying to apply his results to Pereboom's four-case argument. This is because there are aspects of both the study itself and Feltz' conclusions that are cause for concern. One thing we may be concerned about is the methodology. As Feltz acknowledges, Pereboom uses philosophical language that we would not expect the typical lay-person to understand. Although Feltz mitigates this issue as best as possible it is still difficult to say just what effect this has on participants' responses. How do they interpret phrases like, "Bill identifies with the desire to kill Mrs. White and the resulting action," and "the desire to kill Mrs. White is stronger than any competing desire"? For those not familiar with the compatibilist conditions these phrases are intended to reflect it may be thought that Bill's identification with his desire is something distinct from his being manipulated or that Bill somehow has control over which competing desire is strongest. From the lay-perspective these types of phrases can easily be seen as conflicting with the manipulation rather than aspects of Bill's psychology that arise from manipulation. After all, we have already seen from Sripada's study that some participants read into the vignettes, and they may even use these unjustified inferences to make their judgments about freedom and responsibility. If we want to draw conclusions about Pereboom's four-case argument from this study we should be sure that participants understand the extent of the neuroscientists' influence.

Another difficulty with Feltz' methodology relates to how responses to questions were measured. As we've seen, participants were asked to express their agreement with particular statements on a 1-7 scale. Feltz may be right when he says that "the folk seem comfortable with treating moral responsibility as a degree concept" (Feltz 2012, pg. 60), but what are we to make of these responses? The results of both studies indicated that it was only in the determinism cases where the mean response was above 5.5 and there was no case where the mean response was below 3.5 (Feltz 2012, pg. 59, Fig. 1). It is unclear whether we should treat these responses as expressing partial agreement, partial disagreement, or the participants' unwillingness to commit one way or another. Feltz' analysis suggests that we should treat tendencies toward agreement as judgments of freedom and responsibility while tendencies toward disagreement should be taken as opposite judgments. This is useful for drawing conclusions, of course, but it is by no means the only way to interpret the responses. We might think that some people are inclined toward holding Bill responsible but aspects of the manipulation make them unsure that this response is appropriate. Likewise, some people may be inclined to think that the manipulation alleviates Bill of responsibility but the facts about his psychology give them reason to think otherwise. Without prompting the participants in a way that presses them to commit one way or the other, the

meaning of these responses remains ambiguous.

Beyond the methodology, there are even more important aspects of Feltz' study that are cause for concern. When considering a project that examines folk intuitions about Pereboom's argument we must ask ourselves, what *role* do intuitions play? That is, we need to determine which of Pereboom's premises these intuitions are even in a position to support or weaken. Feltz attempts to do this by outlining the general structure of the four-case argument:

> 1. A manipulated agent is not free.
> 2. There is no relevant difference between a manipulated agent and an agent in a deterministic world.
> 3. If there is no difference between a manipulated agent and an agent in a deterministic world, then an agent in a deterministic world is not free.
> 4. Therefore, an agent in a deterministic world is not free. (Feltz 2012, pg. 55)

From here he suggests that the four-case argument is meant to provide evidence for premises 1 and 2 and that folk intuitions can be used to test a prediction made by Pereboom—"If I am right, it will turn out that no relevant difference can be found among these cases that would justify denying responsibility under covert manipulation while affirming it in ordinary deterministic circumstances, and that this would force an incompatibilist conclusion" (Pereboom 2001, pg. 112). The discussion of the results of his study reveals that Feltz believes both premises 1 and 2 are subject to intuitional scrutiny, as is this prediction made by Pereboom.

Now it is certainly the case that premise 1 in the four-case argument is an appeal to our intuitive judgment of manipulated agents. In fact, Pereboom has said that a crucial assumption of his argument is that, "initially it will be agreed, at least provisionally, that the agents in the remote and local manipulation cases are not morally responsible" (Pereboom 2008, pg. 164). It would be a mistake, however, to suppose that either premise 2 or Pereboom's prediction can be addressed by folk intuitions. Determining which differences between a manipulated agent and a determined agent might be relevant to freedom and responsibility is a philosophical endeavor—one that requires argument and justification. Even if some participants go through a process of deep and rational deliberation that leads them to a thoughtful judgment, we could not know it by looking at the level of agreement with particular statements. Moreover, we could not know which judgments, compatibilist or incompatibilist, were arrived at by which means—rational deliberation or gut feelings. This is important

because one of Feltz' conclusions is that his study can be used to support a soft-line reply to Pereboom's argument. This claim relies on the idea that premise 2 of the four-case argument can be analyzed by intuitions since the soft-line reply holds that there are responsibility-relevant differences between the manipulation cases and the determinism case. Since it is not the role of intuitions to tell us which aspects of the cases are responsibility-relevant, the claim that this study supports a soft-line response fails.

Feltz also makes the related claim that, "Pereboom's prediction is just not true: people do find differences between the four cases" (Feltz 2012, pg. 60). As we have seen, Pereboom does indeed say that "no relevant difference can be found among these cases…," however we should not think that this claim is testable by intuitions. Rather, this is a claim that Pereboom both intends to and does support through argument. Moreover, *intuitions* about any of these cases are unlikely to change at all, let alone through mere exposure to them. Pereboom's argument is not an attempt to alter intuitions—his claim is that, upon considering each case, the first two can function as analogies that make "rational the belief that the ordinary determined agent is not morally responsible" (Pereboom 2008, pg. 162). The generalization, combined with arguments that all cases are similar in the relevant ways, is intended to elicit the *belief* that no Plum is morally responsible, even if our contrary intuitions persist. Thus Feltz' claim that "the general pattern of responses…does not support Pereboom's predictions" (Feltz 2012, pg. 53) may be accurate but it does not weaken the four-case argument.

The part of Feltz' study that is most intriguing, and can plausibly have an impact on the four-case argument, is his claim that the results support a hard-line response. Since Pereboom's argument does rely on the initial intuition that manipulation undermines responsibility, the fact that participants tended to rate Bill free and responsible in the Non-Intentional manipulation cases is significant. If we accept that it was the intentionality that undermined freedom and responsibility, and that removing this aspect gave the participants a manipulated agent who satisfied CAS, then it is these compatibilist-friendly judgments which reflect intuitions about a properly manipulated agent. The reason that we should be hesitant to draw this conclusion, though, is a reason that Feltz acknowledges—the intentionality of a manipulator should not be a factor. Since we cannot point to a feature of intentional manipulation that both bears on the responsibility of a manipulated agent and is not present in non-intentional manipulation, we cannot plausibly say that intentionality makes a difference. So while it may *in fact* play a role in how the folk make these judgments, we would be hard-pressed to say that it *ought* to play a role. From here we are left with opposing judgments in two cases, Bill in Intentional

manipulation and Bill in Non-Intentional manipulation, both of which are equivalent except for an irrelevant factor—intentionality. In order to maintain that these results support a hard-line response to the four-case argument Feltz needs to justify preferring the responses to the Non-Intentional case over the Intentional one. This requires an appeal to McKenna's response to the four-case argument, and thus brings us back to the McKenna-Pereboom debate.

4. CLARIFYING CONSIDERATIONS AND CLOSING REMARKS

Given the difficulty inherent in measuring folk intuitions about complex cases, it does not seem likely that experimental philosophy will settle questions surrounding the four-case argument. While Sripada's experiment may demonstrate that folk intuitions are consistent with compatibilist claims, it does not rule out the possibility that ultimate control is a significant factor. Feltz' study reveals that intuitions may support a hard-line reply to Pereboom's argument but to determine whether this is actually the case we must turn our attention *away* from folk intuitions. Perhaps most importantly, these types of studies may be able to tell us what people's intuitions actually are but they won't necessarily tell us what intuitions are *rational* to have. This, it seems, requires the type of debate that has taken place between Pereboom and McKenna.

In response to McKenna's hard-line reply to the four-case argument, Pereboom suggests a clarification of which initial attitude is appropriate to hold toward Plum in Case 4. The appropriate attitude is that of the "neutral inquirer," one who holds that "determinism provides a reason for giving up the responsibility assumption, but claims that so far the issue has not been settled" (Pereboom 2008, pg. 162).[6] A key feature of this attitude is that, although the issue is not *yet* settled, further clarifying considerations may allow the neutral inquirer to be persuaded one way or the other. Pereboom argues that, "adducing an analogy in which one's intuitions are clearer might itself count as the relevant sort of clarifying consideration" (Pereboom 2008, pg. 162). Thus if the neutral inquirer is unsure about Plum's responsibility in Case 4, comparing intuitions about Case 1 and perhaps 2 can make this judgment clearer. Since most agree that the initial intuition is that these manipulated Plums are not responsible, this works in the incompatibilist's favor. It also means that running the generalization backwards does not necessarily result in agnosticism about Case 1. In viewing these cases as clarifying considerations, the neutral inquirer may very well be persuaded of Plum's nonresponsibility upon reaching the

[6] Pereboom contrasts this attitude with the "resolute compatibilist" and the "confirmed agnostic." The former holds that determinism does not even pose a threat to responsibility and thus "enquiry into the issue is closed." The latter is agnostic about the issue but also considers enquiry closed. (Pereboom, 2008, pg.161-2)

covert manipulation cases.

Although McKenna concedes that intuitions about Case 1 and Case 2 can move the neutral inquirer somewhat toward an incompatibilist judgment about Case 4, he does not think that this means victory for the four-case argument. What can be resisted is that, "Cases 1 and 2 are *sufficiently* compelling to move one fully off the neutral inquiring response" (McKenna 2013, pg. 11). In support of this position McKenna offers his own clarifying considerations, ones that would move the neutral inquirer, at the very least, further away from an incompatibilist judgment. First, he suggests calling attention to the fact that properly manipulated agents can be just like ordinary people. By making the psychological features as clear as possible the neutral inquirer can see the manipulated agent as someone who lives a rich moral life rather than "a mere comic book sketch of a full-blooded person" (McKenna 2013, pg. 11).

Next McKenna notes that Pereboom's cases 1 and 2 are not the only cases that count as clarifying considerations. There are also real-life examples of what we could call manipulation which are similar in kind to Pereboom's Case 3. McKenna suggests that, most of the time, "the natural, intuitive response to these kinds of cases is to persist in regarding such agents as free and responsible" (McKenna 2013, pg. 12). McKenna also argues that these real-life cases should be weighed more heavily in the philosophical debate than should Pereboom's Case 1 and 2. This is because there is better reason to trust these intuitions since we have evolved to make judgments in ordinary contexts rather than bizarre ones, and the manipulated Plums are certainly bizarre. Thus McKenna concludes, "other-things-being-equal, intuitive reactions to closer-to-home cases offer a higher degree of reliability given that our conceptual training and modes of performance in these contexts is cultivated and honed in ways that they are not when we apply them in the psychologically bizarre cases" (McKenna 2013, pg. 15). Taking all of this into account, then, it is not clear that Pereboom's clarifying consideration are enough to move the neutral inquirer to an incompatibilist judgment.

We must ask now, where does this leave the four-case argument? If McKenna is right that a neutral inquirer ought not to be driven to an incompatibilist conclusion given *all* the clarifying considerations, then the compatibilist can declare victory. The four-case argument depends on the nonresponsibility of Plum in cases 1 and 2 and, so long as this is not clearly the case, the argument fails. The incompatibilist, however, is not likely to give up this easy. If it can be established that intuitions about bizarre cases are just as trustworthy as those about ordinary cases, or that these types of cases have some other feature that provides important insight, then this may lessen the force of McKenna's argument. Likewise, if it can be shown that the causal processes underlying

Plum's psychology significantly undermine his responsibility *even if* his rich form of compatibilist agency is made clear, then this too would favor Pereboom's argument. Whether either of these counterarguments can be defended is open to debate, though one thing seems clear—any further attempts to advance the four-case argument will require meeting the objections raised by McKenna, and we can expect the incompatibilist to accept this challenge.

BIBLIOGRAPHY

Ayer, A. J. Freedom and Necessity. *Philosophical Essays*. London: Macmillan. 1954

Feltz, Adam. Pereboom and Premises: Asking the Right Questions in the Experimental Philosophy of Free Will. *Consciousness and Cognition*. 22(1): 53-63. 2013

Fischer, John and Ravizza, Mark. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press. 1998

Frankfurt, Harry. Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*. 68(1): 5-20. 1971

McKenna, Michael. A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument. *Philosophy and Phenomenological Research*. 77(1): 142-59. 2008

McKenna, Michael. Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin. *Philosophy and Phenomenological Research*. 87(3): Early View. 2013

Mele, Alfred. *Free Will and Luck*. Oxford: Oxford University Press. 2006

Pereboom, Derk. A Hard-Line Reply to the Multiple-Case Manipulation Argument. *Philosophy and Phenomenological Research*. 77(1): 160-70. 2008

Pereboom, Derk. *Living without Free Will*. Cambridge: Cambridge University Press. 2001

Sripada, Chandra Sekhar. What Makes a Manipulated Agent Unfree? *Philosophy and Phenomenological Research*. 85(3): 563-93. 2012

Wallace, R J. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press. 1994