Ex Nihilo

The Undergraduate Philosophy Journal of
The University of Texas at Austin

Acknowledgements

It is difficult to offer but thanks for the tirelessness and dedication of those with whom you have worked so closely for so long. But, as nothing in the production of this journal has not been fraught with difficulty, I shall take the task as I find it.

This year's *Ex Nihilo* was a joint effort with the Texas Undergraduate Philosophy Conference. The conference's coordinator, Julie Hunter, is as much responsible for the former as I, and even more so responsible for the good one may find herein. Her organizational skills, patience, devotion, attention to detail, and willingness to take on neglected responsibilities served both as an asset and a matchless example to the rest of the staff and organizers. I thank her first.

Little in UT's Philosophy Department is accomplished without my old friend and colleague, Jeff Casey. Whether it is due to his overbearing personality or his ceaseless dedication is difficult to say—but indeed I suspect the latter. Jeff provided an otherwise inexperienced group of undergraduates with the knowledge, confidence, and ability that come with his advanced distinction and age. He deserves my thanks as well.

I should also like to thank the department chair, Nicholas Asher, for his support; the *Ex Nihilo*/Texas Undergraduate Philosophy Conference submissions committee; the conference participants, respondents, and attendees; our keynote speaker, David Sosa; the philosophy department staff; and, not to be forgotten, Justin Cone for his rather professional fliers.

Further, let me extend my personal gratitude to that disturbingly committed group of individuals, known officially as the Undergraduate Philosophy Association, for turning the loose conglomeration of undergraduate philosophers in our department into a close-knit community. *Ex Nihilo* is but one of the UPA's many projects, including the Friday Lunch Lecture Series, the Analytic Philosophy Reading Group, and the Texas Undergraduate Philosophy Conference.

When philosophy is done neither in the classroom nor the conference room, it is done on the written page. I am not ashamed to offer you what follows; save for one paper, I may say with a measure of objectivity that it is solid philosophy. Enjoy.

MICHAEL JOHNSON
Editor

Contents

On the Supposed Duty of Humanitarian Intervention

Recent U.S. military humanitarian interventions were often preceded by a strong moral outcry by journalists and the international community. Indeed, during the Kosovo war, President Clinton framed the issue in terms of moral imperatives in a Memorial Day speech where he said, "Our objectives in Kosovo are clear and consistent, both with the moral imperative of reversing ethnic cleansing and killing, and our overwhelming national interest in a peaceful, undivided Europe."[1] Clearly, the President felt it necessary to claim that America not only had a national security interest in Kosovo, but that there was also a strong moral imperative to intervene. This belief that military intervention is justified by a moral imperative is so prevalent that at the start of the recent Liberian civil war, Liberian nationals stacked bodies of their dead outside of the U.S. embassy.[2] The purpose of this was to show that America had an obligation to intervene in the bloody civil war. This essay will examine the putative moral obligation to intervene militarily in humanitarian crises and argue that because of the nature of moral obligations and the nature of the state, it is impossible that the United States ever has a moral obligation to pursue

---

[1] McQuillan, 1999.
[2] Paye-Layleh, 2003.

humanitarian interventions.[3] I will then examine some objections to this view.

The Nature of Moral Obligations

To claim that someone has a moral obligation to do x is to claim that one has a *duty* to do x. That is, one should and must do x. Is this what people really mean when they claim that the United States has a moral obligation to intervene in nation x? Do they mean that the corporate entity "the United States" has a moral duty, or do they mean that the aggregate of United States citizens has a moral duty? And: a duty to whom? These questions are rarely asked in the literature on humanitarian interventions and almost never answered. If we are to make sense of the statements generally made by pundits and philosophers about the need to intervene, we must look at these questions seriously and come to some solutions as to what it means to say that the United States has a moral obligation to intervene in country x.

First, because a humanitarian intervention is usually the use of force against one group or another, the action requires the special justification that is necessary for any use of military force. The justification for a humanitarian intervention cannot be self-defense in the traditional sense, because the nation that is using force is doing so on behalf of another group, not itself. Therefore, not only do we need to ground the moral obligation, we also need to justify military intervention on behalf of others generally. It is often claimed that humanitarian intervention is justified on self-defense grounds. A third party, the intervener, is merely prosecuting the self-defense claim for the attacked party. The analogy here is to the state prosecuting a crime on behalf of the victim. This seems prima facie reasonable and harkens back to Walzer's "domestic analogy" in the realm of international law.[4] If we accept the "domestic analogy" concerning the prosecution of others' claims to self-defense, then it seems that there is no special problem

---

[3]It should be noted that when I refer to the "state" I am using this term to mean a liberal democratic state, or some form of liberal representative state. To substitute an absolute monarchy or a dictatorship would defeat many of my arguments, though it seems that, for the most part, the debate on humanitarian interventions is based in liberal states.

[4]Walzer, 2000, p. 58.

in justifying a humanitarian intervention as an acceptable possibility
of action.

The more serious issue is how to justify humanitarian interventions
as an obligation; that is, how to find the *duty* for one nation to pros-
ecute the self-defense claims of another nation. One way to unpack
the often-vague arguments about moral *duty* to intervene is in the
terminology of human rights. Nigel Dower explains it in this way:

> If we consider 'human rights' justifications in these terms,
> it is clearly linked to justice and to the making of rights
> claims. It is odd to separate humanitarianism from human
> rights, since often the justification for humanitarianism is
> in terms of human rights.[5]

The argument above seems to be that humanitarianism is linked to
human rights claims, and that the purpose and warrant of an inter-
vention is to prevent or stop human rights violations. Clearly, every
person has a duty not to violate certain rights of others in civil society.
Stealing is violating a right to property; murder a right to life, etc. It
seems plausible that this could be applied to the international realm
in the same way that it is applied to the domestic realm.

Several problems with this theory make this leap difficult, however.
First, human rights are contradictorily defined in the international
community, making it difficult to adjudicate disputed rights. The best
example of this is the UN declaration of human rights. The preamble
to the declaration claims that the list is the codification of the basic
human rights that everyone possesses regardless of nationality. The
most obvious contradiction of rights in the document is the contradic-
tion between Article 17, which states that, "Everyone has the right to
own property alone as well as in association with others," and Articles
22-26.[6] The latter articles claim that it is a fundamental human right
to enjoy "social security," "protection against unemployment," and,
the most infamous of human rights, the right to "periodic holidays
with pay."[7] It is unclear how these welfare rights are to be enforced
other than by confiscating the property of one and giving it to another,

---

[5]Dower, 2002, p. 75.

[6]United Nations, *Universal Declaration of Human Rights* (New York: UN,
1948).

[7]Ibid.

contradicting the right to property in Article 17. Further, are we to give the same weight to violations of the right in Article 24, the right to paid holidays, as we give to violations of Article 3, the right to life? Clearly we are not; though this particularly muddled codification of rights does not give us any way to decide between rights claims. Tom Palmer makes this problem clear with an example:

> Let us say that Bill needs medical care and necessary social services and that Janet is a doctor. If Bill has the right to Janet's services (Article 25), but Janet has the right to rest, leisure, reasonable working hours, and periodic holidays with pay (Article 24), and those conflict (as they surely do on occasions), whose rights are to be realized?[8]

Surely, a nation such as the United States is not meant to invade another country because not everyone has a right to paid holidays. This example shows the poverty of one of the most central international human rights codes, making justification of humanitarian interventions on the grounds of the UN declaration nearly impossible.

It may be possible, however, to narrow the usage of rights down to a very limited number, thereby giving us certain situations that would warrant intervention. Genocide, the mass extermination of a certain group, is an example of this type of rights violation. Even this, though, is somewhat confusing. Genocide is just the mass, systematic killing of a group of people, making it no different in the realm of duty from the unjust killing of one person. The important point in regard to rights and duties is not the amount of rights violated, but rather the type. While it surely makes a difference whether 5,000 or 500,000 are killed unjustly, if there is a strict or perfect duty to intervene in the self-defense of others, it should apply regardless of the number killed. This duty seems to prove too much, though. A nation cannot be expected or allowed to intervene every time someone is murdered, not even in the case of several hundred murders. Clearly, the amount of people killed in genocide has some bearing on the duty to intervene, though this is not true in the case of a strict duty to aid in the self-defense of others.

A perfect duty to aid in the self-defense of others may be too strict. Every nation on earth would be required to intervene in almost every

---

[8]Palmer, 2001, p. 47.

other nation to prevent some abuse going on there. This is not only unworkable in theory—it is also somewhat incoherent. Some have tried to solve this problem by claiming that the duty to aid in the self-defense of others is an imperfect duty rather than a perfect duty. The distinction between duties is described as:

> A perfect duty is one for which there is a corresponding right. For example, if I have a duty not to execute prisoners of war, you, as a prisoner of war, have a right not to be executed. An imperfect duty is one for which there is no corresponding right.[9]

An example of an imperfect duty would be charity. I may have a duty to give charity, though no one has a right to my charity. No one in particular can claim that any of my wealth belongs to him based on his right of charity. Imperfect duties are like virtues: they are not directed at anyone in particular and they do not create rules for action, merely guidelines. Writing about this in regard to humanitarian interventions, Holzgrefe claims that, "States have a discretionary right to intervene on behalf of the oppressed...States may discharge it at their own discretion and in the manner of their own choosing. The victims of genocide, mass murder, and slavery possess no 'right of humanitarian rescue'—no moral claim to the help of any specific state."[10]

   If, therefore, the moral duty to intervene to stop human rights abuses is merely an imperfect duty, the moral obligation seems to collapse to something much weaker than is normally asserted. To say that a state has a "moral obligation" to intervene in defense of another group is merely to say, "all things being equal, it would be better if some state intervened to help this group than not." This is not a very strong claim, and it can be weakened even further. The *ceteris paribus* clause at the beginning of this moral claim means that many other factors must be evaluated to determine the extent of the moral duty. If the cost to the intervening state's citizens is too high, or the benefit of intervention is too low, the moral claim could be made untenable. The cost of intervention may have another aspect: that is,

---

[9]Holzgrefe, 2003.
[10]Ibid., pp. 26-27.

the cost of contradicting or putting into tension other perfect duties that the state may have to its own citizens. Allen Buchanan explains:

> There is no suggestion that the state must do anything to serve the cause of justice in the world at large. What makes the government legitimate is that it acts as a faithful agent of its own citizens. And to that extent, government acts legitimately only when it occupies itself exclusively with the interests of the citizens of which it is the government.[11]

The state may have previous contractual duties that it must satisfy before it can dispose of its imperfect duties to justice in general. It seems also that the duties, being contractual in nature, are perfect duties. Further, the benefits of a given intervention are often unclear and hard to quantify, even after the intervention is complete; because of this, the obligation to humanitarian intervention, even in its imperfect form, seems little more than a guidepost.

The moral obligation to participate in humanitarian interventions is clearly an imperfect rather than a perfect duty. As an imperfect duty, it is further qualified by a clause regarding costs and benefits. The duty to promote justice abroad seems more similar to a duty to help the poor than a duty to keep promises; however, just because no strong duty exists, this does not mean there is no reason for doing something. The analogy here is to Thomson's "Minimally Decent Samaritan."[12] We might say that if the cost to a humanitarian intervention is very low and the expected success is high, a "Minimally Decent Samaritan"–type state should intervene. This still does not claim that the state has a perfect duty to intervene, and further, it does not override any conflicts the imperfect duty to intervene may have with other duties, but it does provide a case in which it is plausible that a state could be required on some moral level to intervene in another country.

Although there is not much of a duty to humanitarian intervention left, there is still the question as to what it means to claim that a state has any duties at all. If states cannot possess moral duties at all, it makes no difference what type of duty there is.

---

[11]Buchanan, 1999, pp. 74-75.
[12]Thomson, 1971, p. 62.

The Nature of the State as an Actor

The first question that must be answered when there is a claim of a moral imperative is, "a moral imperative for whom?" In the case of the Liberian situation, the answer was obviously the United States. The problem that needs to be addressed, though, is whether the United States is the type of thing that can possess moral obligations. If the state cannot possess a moral duty by its nature, it is a mistake to talk about moral obligations for intervention. It will be my argument that states cannot possess moral duties and, therefore, that there can be no moral duty to humanitarian intervention.

First, we should look at the nature of the state. What is this thing we call a "state"? One possibility is that the state is a type of person or organism. This would be similar to the way that a limited liability corporation is considered a legal person, which is similar in some ways to a real person. What are the defining features of a person—what are the features that would make us believe that the state resembled a person in a significant way? The defining characteristic of human beings seems to be something like rationality or rational autonomy. While there may be many different interpretations of what these two characteristics are, on some level at least one of the above features seems to be necessary for moral action. The state cannot be said to possess either of these two characteristics. It is even somewhat confusing to think of the state in these terms. The state is made up of people; it is not a person in itself. Although in our everyday language we may say that the state did such and such, this statement is meaningless unless it is shorthand for the claim that a certain collection of people did various things amounting to the effect that such and such. It seems safe then to agree with H.J. McCloskey when he claims, "That the state is not literally a person is so evident that we need not consider such a possible contention further."[13] McCloskey's claim can be taken further, though. It is not only evident that the state is not "literally" a person, but also that it is in no way related to a person.

Although we have seen that the state is not a person, it could still be a considered an "actor" in a non-person sense. The state is made up of many, often seemingly unrelated, people and a number

---

[13]McCloskey, 1963, p. 319.

of complex legal relationships. The fact that the state is reducible to
other things, however, does not prove that the state lacks the requisite
unity to be an "actor." Many things, if not most things, are reducible
to their constituent parts. Even people can be reduced to their organs
and even to individual cells and atoms. What needs to be proven is
that states differ in significant ways from things that are generally
said to have moral duties. For instance, we do not say that a bear acts
"wrongly" when it attacks a person, or that our computer is "evil" for
crashing. It does not make sense to talk about computers or bears in
moral language because we do not consider them to be the types of
things that can act morally or immorally. People, however, are said
to act "wrongly" or to be "evil." The state seems to be more similar
to a computer than to a person in this respect. One problem with
this view is, however, that we generally do not think it is odd to say
the "state acted wrongly." This phrase seems perfectly acceptable in
common English.

Why is it odd to say that a bear or a computer is "evil," but not
a person? The reason that it seems odd to claim that a bear is "evil"
is because we generally consider an absolute necessary condition of
moral or immoral behavior to be the capacity for autonomous action.
A computer is clearly not autonomous; that is, it is not self-ordering
and it does not make its own laws. A computer requires the user to
direct its every action, defeating the necessary condition of autonomy.
We also, for the most part, do not consider a bear to have the rational
faculty that allows it to be self-directing in the moral sense. We would
claim that a bear is controlled wholly by instinct or something similar,
though this is a more controversial claim.[14] For a state to be said to
have moral duties, we must claim that a state is autonomous in the
same sense that a person is.

On face, however, a state is not autonomous. States literally do

---

[14]It is odd, though, that most who claim that animals possess rights do not
make the corollary claim that they also possess duties. For instance, if bear x has
the right against being killed by person y, person y has a duty not to kill bear x.
Most proponents of animal rights do not make the further claim that if person y
has a right not to be killed, bear x has a duty not to kill person y. Even if animals
did have moral duties, they clearly do not have the requisite legal duties. It is
unclear what a legal duty for an animal would even mean, would we take bear x to
court if it murdered person y? What procedural rights would bear x have? These
questions generally go unanswered.

not make their own laws. This is most clear in the case of the United States, where there is a written constitution that is the law of the state. This law was not created by the state; it was prior to the state, yet it binds the state. Further, in the case of ordinary laws, these are made not by the state per se but rather by representatives of the people. One could argue that since the government of a democracy represents and derives its power from the people, the government is nothing more than the people. This claim is incoherent, though. If there were a revolution, the people would not be rebelling against themselves: they would be rebelling against the state. The state is a set of contractual relationships made by the people; however, in this context, it is distinct from the people. Further, the state is not even the same thing as the people who make up the government. The state cannot be the same thing as the president or a certain legislator. It cannot even be the same thing as the aggregate of all government employees, because the employees are not only constantly changing in person, they are also constantly changing in responsibility. These individuals occasionally act in concert, though not always. The state, therefore, being neither any individual person nor group of people, can be said to be only the name of a complex set of contractual relationships. This is an odd and counterintuitive way of thinking of the state, but it is the only way that does not devolve into incoherence.

If the state is nothing more than a set of contractual relationships, to what extent can the state be said to exist at all? It exists in a similar, though not in the same way, that a limited liability corporation or a contractual trust exists. The example of the trust is better than the example of a limited liability corporation because the corporation requires the charter of the state to give it certain privileges, whereas in most cases, a trust can be entered into purely by private contract. A private trust is an entity created when one person contracts with another about their property ownership rights. The trust exists to fulfill the contractual obligations that the trustees have created. In this sense, a trust is a kind of "obligation machine": that is, it operates only to fulfill previously created and emergent contractual obligations. The trust as "obligation machine" is not a moral actor, however. It has no autonomy and is purely a legal fiction, or more clearly a legal metaphor. The same is true of the state.

Under all liberal theories of the state, the state is also nothing more

than an "obligation machine."[15] Citizens ceding certain enforcement powers to the state form the state. No rights are ceded, however, and neither are any duties.[16] This means that the state is an entity that possesses only powers and disabilities, not rights and duties. This is obviously very different from a person, who possesses not only legal rights and duties, but also moral rights and duties. It is not proper to speak of the state not having the "right" to do x, but rather of it not having the "power" to do x.

Therefore, if the state is a power-given, though right-less, entity, it becomes clear that the claim that the state can have moral obligations is false. In the same way that no citizen can have a moral obligation to the state, the state has no moral obligations to anything or anyone. The state is not a moral actor, nor could it be one. Of course, individuals have the same moral duties and rights when they act under the aegis of the state that they have in civil society. So while it may be improper to claim that the United States acted immorally when it bombed Hiroshima, it would be perfectly proper to claim that Harry Truman, or the bomber pilot, or any number of other people acted immorally.

To summarize, the argument above is that because the state is not a person and does not have the capacity for autonomous action, the state cannot be considered a moral actor. That is, the state can act neither rightly nor wrongly in the moral sense. The state is an "extra-moral" actor, which has neither moral obligations nor moral rights. It is therefore meaningless to claim that there is a moral obligation for a state to perform a humanitarian intervention, even in the imperfect sense of moral obligation presented above. The issue is not completely settled, however, because there are still some lines of attack open to one

---

[15]By "all liberal theories," I am referring to all liberal versions of how the state emerged from the state of nature. Under Locke and Nozick as well as Rawls and Hobbes, there is some type of contract, real or ideal, that forms the state to fulfill certain obligations. In the case of Nozick, however, this contract may be tacit.

[16]This may seem like an odd statement because political theorists often speak of the "rights of the state" or the "rights we gave up when we entered the state." This is just sloppy terminology, however, and what these theorists apparently mean is "power" not "right." For a detailed analysis of the difference of a "right" and a "power" see: Wesley Newcomb Hohfeld 1919: *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, ed. by Walter Wheeler Cook. New Haven: Yale University Press.

who would disagree with this presentation and to those who, despite their acceptance of the arguments above, may think there is still an obligation to perform humanitarian interventions.

Further Objections

Even if one thought the argumentation above was sound, there is still a way to sneak in a moral obligation to humanitarian intervention through the back door. One could argue that the obligation to intervene for humanitarian purposes is not a duty that falls on the state, but rather a duty that is held by the members of a state. In this way, there could still be a moral obligation that groups of people in a given geographical area have to people in another geographical area, regardless of whether or not the state is a moral actor.

It is not my purpose here to argue against humanitarian intervention categorically; however, this objection requires a response, so as not to make the argument above fall into triviality. One argument against this objection is to claim that, at best, any obligation to prosecute humanitarian interventions that is held by the citizens of a certain state is merely an imperfect obligation. Further, because "ought" implies "can," even if there is an imperfect obligation of the members of a state to stop human rights abuses in other states, their ability to do so is, in most cases, severely limited, reducing the moral claim to almost nothing.[17] This would also mean that everyone in the world has a duty to prevent or stop humanitarian abuses in other countries, including citizens of the state in which the abuse is being committed. If everyone in the world has the same moral duties to a state where human rights abuses are being committed, the issue becomes one of efficacy, not morality. Which people are best able to prevent or stop an abuse? This is a difficult issue to decide, though not a moral issue.

Along these same lines it may be claimed that because every individual has a duty to prevent or stop human rights abuses, albeit an imperfect one, individuals may cede their duties to the state so

---

[17]By "severely limited" I mean that most groups of people outside of the state apparatus do not have the requisite funds or equipment to intervene militarily in another nation. They could, however, take other action to alleviate the harm by sending aid, etc.

that the state now has a moral obligation to rectify a humanitarian abuse. This is not an acceptable move, however, because of my above analysis of the state, which makes it impossible for the state to have any moral duties. Although the state cannot have any moral duties, it may be possible for citizens to give the state a power, or allow it to use a power it already has, to rectify humanitarian abuses for the purpose of fulfilling their imperfect duties. This could be done, however, only if this power was not prevented by a previous disability that the state already had. In the case of the United States, it would be very difficult to institute this type of system legally, as many disabilities prevent the state from acting this way. Ultimately, though, even if one did take this line of argument, it would not defeat the claim that the state cannot have moral duties, which is the crux of the argument above.

Another objection could be that when one claims that the United States has a moral obligation to intervene in country x, what is really meant is that the citizens of the United States have a duty to intervene in country x. This is similar to the previous objection, though less powerful. First, "the United States" does not mean the aggregate of people in the United States. One reason for this is that if it did mean that, the phrase "the United States invaded Iraq" would mean that all the people in the United States invaded Iraq. Not only has this never happened, though the "United States" has invaded Iraq, but it is implausible that anyone ever has this sense of "United States" in mind when one claims such things. In the proposition *the United States has a moral obligation to intervene in country x*, "United States" is used in the same sense as the proposition "The United States invaded Iraq;" so if the latter does not mean the aggregate of people in the United States, neither does the former. It is further implausible to maintain that individual citizens of the United States or any other country have a duty to rectify humanitarian abuses, as even the might of the United States military is often unable to do so.

The State of Humanitarian Interventions

Although the state has no moral obligation to perform humanitarian interventions, nor could it have such obligations, this does not com-

pletely rule out humanitarian interventions. It may be possible for liberal states to make provisions for humanitarian interventions. It is also possible that non-governmental organizations may perform humanitarian projects, though at least at the present, there are very few NGO's that would be willing or able to perform military interventions. The Red Cross and other groups, however, could assist groups in need and individuals could fulfill their imperfect obligations by donating to these groups or volunteering to help them. Indeed, it may be that these NGO's can have a more positive effect than state-based military interventions. The United States and the U.N. have a spotty record, at best, of actually solving problems and, all too often, it seems that a humanitarian intervention is merely a hiatus for genocide or civil war, rather than a resolution of it.

Also, because of this spotty record, it may always be in the interest of citizens of a state to prevent or severely limit the ability of that state to participate in humanitarian interventions. In one sense, humanitarian intervention is a perfect example of a public good. This has led some to believe that the state is required to supply it; though just because something is a public good does not mean the government is required to supply it. It does mean, however, that the good will generally be oversupplied or inefficiently supplied unless the costs of the good are internalized. At the moment, the public good of humanitarian intervention to the world is being internalized primarily by the United States, leading to calls for more of the good that do not reflect its efficient level and cost.

Further, in the realm of foreign policy, the primary contractual obligation of the officials of the state is to protect their citizens. If the pursuit of humanitarian missions makes this primary obligation harder to fulfill, it is acting outside of its power. Of course, it is hard to quantify what constitutes "too much" humanitarian intervention, or what level of humanitarian intervention makes the citizens of a state unsafe; however, given the potential threats on the international scene, a state should be able to make a rough approximation of when this threshold is met. It is also possible that in the pursuit of national security—that is, national protection—humanitarian interests could be realized. One could argue not only that it was right for the United States to intervene in Kosovo on security grounds, as President Clinton did, but also that a humanitarian objective was secured. It is important, however,

not to concoct specious security justifications for humanitarian goals.

After 9/11, we realized that the post-cold-war world is not the "end of history." The United States is not invincible, and the security of her citizens is the primary goal of the officials of the state. This does not rule out humanitarian interventions altogether; though I would suggest that a policy of "intervener of last resort" is the proper policy to take in lieu of the lack of moral obligation and the cost of intervention. Often people claim that the state should "do something" to solve some problem. This is a dangerous sentiment, however, because it combines activism with vagueness. It is rarely clear who should "do" and what the "something" is that should be done. As heartless as it may seem, often the best policy is to do nothing. It is up to the individual to decide what situations constitute sufficient warrant to "do something" and what that "something" should be.

<div style="text-align: right;">

JOHN THRASHER
University of Colorado, Boulder

</div>

Bibliography

[1] A. Buchanan. The Internal Legitimacy of Humanitarian Intervention. *Journal of Political Philosophy*, 7:71–87, 1999.

[2] N. Dower. Violent Humanitarianism—An Oxymoron? In A. Moseley and R. Norman, editors, *Human Rights and Military Intervention*. Ashgate, Burlington, 2002.

[3] W. N. Hohfeld. *Fundamental Legal Conceptions as Applied in Judicial Reasoning*. Yale University Press, New Haven, 1919.

[4] J. L. Holzgrefe. The Humanitarian Intervention Debate. In J. L. Holzgrefe and R. O. Keohane, editors, *Humanitarian Intervention: Ethical, Legal, and Political Dilemmas*. Cambridge University Press, Cambridge, 2003.

[5] H. J. McCloskey. The State as an Organism, as a Person, and as an End in Itself. *The Philosophical Review*, 72:306–326, 1963.

[6] L. McQuillan. Kosovo a 'Big Test' of Moral Values, Clinton Says. *Reuters*, May 31, 1999.

[7] T. G. Palmer. Saving Rights Theory from Its Friends. In T. R. Machan, editor, *Individual Rights Reconsidered: Are the Truths of the U.S. Declaration of Independence Lasting?* Hoover Institution Press, Stanford, 2001.

[8] J. Paye-Layleh. Liberians Blame U.S. as Carnage in Capital from Rebel Attacks Continues for Third Day: Bodies of Children Laid at U.S. Embassy. *The Gazette*, June 27, 2003. Montreal.

[9] J. J. Thomson. A Defense of Abortion. *Philosophy and Public Affairs*, 1:47–66, 1971.

[10] M. Walzer. *Just and Unjust Wars: A Moral Argument with Historical Illustrations.* Basic Books, New York, third edition, 2000.

On the Logical Form of Indirect Discourse

In his essay "On Saying That," Donald Davidson proposes an analysis of the logical form of propositional attitude ascriptions that has been well respected, if not, however, very much accepted, in the philosophical community. The theory is this: statements of the form 'John said that Mary is pretty' are of the logical form 'John said *that*. Mary is pretty,' where 'that' refers to the following utterance, and that utterance, 'Mary is pretty' has the same meaning, and its terms the same referents, as when used at any other time. There are many appeals of this theory. First, it solves Frege's problem, that co-referential terms should be substitutable everywhere *salva veritate*, and yet they are not in indirect discourse. On Davidson's account, such terms are substitutable preserving truth-value of the sentence in which they occur; but it is the introducing sentence, the 'John said that' type, that changes in truth value. Once an element in the following utterance has been changed, it may have the same truth-value, but it no longer is the same utterance, and the referent of 'that' is changed. The second appealing feature of the theory is that it posits "sentences in indirect discourse, as it happens, wear their logical form on their sleeves (except for one small point)" (106). If this were the end of the matter, life would be great. But, as we shall see, it will become impossible for Davidson to keep his theory without revision and that revision, in the end, will have lost the appeal of the original.

Tyler Burge in his "On Davidson's 'Saying That'" offers three objections to Davidson, none of which he believes to be fatal to the

Davidsonian project, only to the paratactic account. Burge's first objection concerns the interpretation of demonstrative 'that's in indirect discourse. He argues that for a sentence such as the first here: 'Galileo said that. That moves.'—where the second 'that' refers to the earth—to be true, Galileo must have referred to the earth in some utterance of his, attributing to it the property of movement. Burge, I think, is on the right track here, but has the wrong principle. Schiffer, in making the same point Burge is trying to make, states the principle thus: "If the occurrence of the singular term t in 'So-and-so said that ...t...' is primary and refers to x, then that sentence is true only if so-and-so also referred to x" (132). It is not that demonstratives per se must co-refer in and out of indirect discourse, but that demonstratives occurring primarily must do so. Schiffer tells us that primary occurrences of terms are those "not properly contained within the occurrence of other singular terms" (132). This point is, nonetheless, incidental. Once we have such a principle stated, the argument follows fairly straightforwardly. Consider: 'Galileo said that. Copernicus said that. The earth moves.' If the second 'that' in this sequence refers to my utterance of 'the earth moves' and it is a primary occurrence of a singular term (what else could it be?) then Galileo must have made some utterance in which he referenced *my* utterance! But Galileo is not Nostradamus, and it is very unlikely he knew either that I would exist or, existing, what I would say.

Schiffer thinks this line of reasoning pushes Davidson into a theory involving utterance-kinds: "the idea," he says, "is to construe 'that' as really having a secondary occurrence in the implicit singular term 'the kind of utterances having the same content as that'" (133). Schiffer thinks that he then has a decisive argument against such a position. But Burge sees a second out for Davidson, in the phenomenon of self-reference. He says, "Where self-reference is more important to the point of a discourse than what is referred to, self-reference is preserved" (194). He actually gives but one example in this paper (though he cites a paper where he supposedly delivers more): the sentence '(3) is not a theorem' where '(3)' refers to the sentence that contains it. We are told, "A translation into Polish would provide a sentence that referred to itself, not to our sentence." But it is unclear that translation and indirect discourse are acts of the same sort. It would seem odd if someone reported: "Michael said that. This utterance of mine is not a

theorem." I may have implied that *the kind of utterances* having the same content as that utterance of the reporters are not theorems, but I certainly didn't mention them; and it seems false to represent me as having done so, even to preserve self-reference.

All the same, though Burge poorly motivates his revision to Davidson, it doesn't mean his revision is incorrect; let us consider it, then. He wants an analysis of a sentence like 'Galileo said that the earth moves' to come out as "([some]x)(Galileo's utterance x and that (next) utterance of mine, taken in the context of this very utterance, make us samesayers). The earth moves." The doubly-embedded context is supposed to operate likewise. I must admit I find this line of reasoning obscure. What *is* the "context of this [that?] very utterance"? Burge calls it "*the interpretative-grammatical context of the introducing utterance*" (196). If what he means by 'context' is "an assignment to indexicals" then the solution works like this: Galileo does indeed say something to the following effect: "there is some utterance of Copernicus' that samesays with that next utterance of mine [i.e. Galileo's]. The earth moves." But I think this type of context shifting only occurs because we slip a hidden quotation into our analysis and ask of our context that it determine who said *it*. We just don't use indexicals that way in indirect discourse. But if this was what Burge meant by context, why the reference to self-reference? I must admit, I am at a loss to make sense of this solution.

Burge believes this next objection to be his strongest one. Clearly, the following argument is formally valid: "Galileo said that the earth moves; so Galileo said that the earth moves." But on Davidson's account the argument looks like this: "Galileo said that. The earth moves; so Galileo said that. The earth moves." Since the two 'that's in the argument refer to two different utterance tokens, the premise and the conclusion of the argument make different claims. Burge says, "to explain the validity [Davidson] would have to enter further premises about the relation between the supposed demonstrative 'that', or an extra premise about its referents" (203). Thus the formal character of the argument's validity is not captured on the account. Furthermore, Burge thinks, Davidson's account "not only fails to represent the formal character of valid arguments in indirect discourse [but even] fails to account for their validity" (204). This is because we can have a "context of evaluation" where the premise is true but the conclusion

is false, namely, the one where the second utterance of 'The earth moves' does not exist. I don't know if this last objection is entirely fair: surely if Davidson must defend the position that such arguments are not formally valid, he still has the claim that only such arguments as have both utterances sound (in any way) valid, and that is all he must then need to account for. In any case, we've come to the point of Burge's article, namely that Davidson's account must at the very least involve abstractions such as utterance-kinds, propositions, etc. Burge goes even farther in his introduction, claiming

> Our grip on the syntax of the relevant discourse is too fundamental and firm for us to be persuaded that we have been misled regarding how to categorize the singular terms or how to divide among the sentences. (191)

However, he does not go beyond citing his intuitions.

But is there not a reply for Davidson? Surely this seems formally valid: Mary is pretty; therefore, Mary is pretty. But it isn't. The only formally valid line of argument in a similar direction would be: Mary is pretty; therefore, Mary was pretty when I made my preceding utterance. Verbs are tensed, and thus indexical. But nonetheless arguments of the form Mary is pretty; therefore, Mary is pretty sound formal if any argument does. It is that in natural language we have a broader notion of "formally valid" than that of logic, and I can't see that Burge's objection is to Davidson's position, so much as it is to natural language itself. Here's another argument that "sounds" formal (as though that made sense): That (pointing to an apple) is red; therefore that (pointing to same object) is red.

Though the two objections of Burge's considered above seemed lacking, he has yet a third one that I believe to be his best. He points out, "we may say that Galileo said something at 6 p.m. 3/15/1683 in the indirect-discourse sense of 'said', without being committed to there being an actual utterance token of ours that samesays Galileo's utterance" (197). If this were his only point in this vein, I think Davidson would be in the clear: for there is nothing about the quantified sentence which to me indicates that 'said' is used to introduce indirect discourse. It seems to be true iff there is some utterance of Galileo's—that is, 'something' quantifies over quotations, or possible-quotations. But the problem becomes stickier when we quantify over

beliefs. Consider: 'Galileo believed something the content of which no-one ever uttered.' Surely this is possible; but Davidson would have it as a contradiction: if believing is a relation between persons and utterances, to have a belief is just to stand in that relation to an actual, concrete utterance. It seems Davidson must concede that the 'that's in propositional attitude ascriptions refer to utterance-kinds, such as need not be instantiated. If we accept that Davidson's account now has to incorporate utterance-kinds, it is time to turn our attention to Davidson's other objector, Schiffer.

Schiffer is a somewhat more dangerous man than Tyler Burge, in that his goal is not for Davidson to discover the magic of utterance-kinds, but rather, in defeating Davidson, to show "that natural languages do not have compositional semantics" (111)—which flies in the face of just about everyone, myself included. Schiffer is going to take up where Burge left off: though Davidson has had to retreat to utterance-kinds, Burge still considered his position a viable one; Schiffer will come down hard on the utterance-kind revision being entertained for Davidson. In the end, he will try to conclude that no revision, for Davidson or any other extensionalist, will be able to handle the problems he raises.

Schiffer comes out of the gate hitting hard. First, he says, "utterance kinds are universals; if one has them, one has properties, propositions and the lot, and no need to worry then about achieving an extensionalist semantics" (130). Davidson's Tarskian program goes down in flames. What's more, says Schiffer, "The logical form of 'Galileo believed that'...should be represented as, say, 'B(Galileo, the utterance-kind to which that belongs)'" (130). So much for Davidson's cute "sentences in indirect discourse wear their logical forms on their sleeves" (106). But even this is not dire enough for Davidson's project to make Schiffer comfortable. He notes that the most likely method for distinguishing utterance-kinds is "by the contents of the utterances they subsume" (131). Schiffer's objection is best put in his own words:

> This is not the tired old objection that Davidson's account
> of saying-that relies on an unexplicated notion of sameness
> of content. *That* objection is a bad one...On the required
> revision, 'that' may still be seen as referring to an actual

utterance, but now it will no longer enjoy a primary oc-
currence. Its occurrence will be ensconced in an implicit
occurrence of the singular term 'the kind of utterances hav-
ing the same content as that', and my point is that we will
not know the *reference* of that singular term until we know
what notion of content is here intended. (131)

This should not be of much trouble to Davidson, as he has a very
clear notion of content, deriving from his theory of meaning. Schiffer,
however, believes that this is not enough.

Schiffer's final objection is intended to show not only that David-
son needs to explicate a notion of content, but that the notion needs
to be embedded in the *logical form* Davidson gives to propositional
attitudes. The objection is this: it is intuitive that to understand
'John said that Mary is pretty' we must know *what John said*; but,
according to Davidson, I can understand 'John said that,' where 'that'
points to a later utterance of 'Mary is pretty,' without understanding
the utterance 'Mary is pretty'—all I have to do is recognize *that* it
is an utterance, and *that* it has the same content as something John
said. The only way for Davidson to save his theory now is to build
the content of the referent of 'that' into the logical form, so that to
understand 'John said that'—'that' referring to 'Mary is pretty'—we
must (1) recognize that it is an utterance, (2) that it has the same con-
tent as something John said, and (3) that the "content determining
feature" of what John said is X. The revision Schiffer then proposes
for Davidson is to treat

(1) John says that Mary is pretty.

as something like

(2) John says the kind of utterances that are X, like that.
Mary is pretty.

The objection Schiffer then raises is that the kind of X's floating
around Davidson's theory are not accessible to average speakers. This
is because Davidson is interested to give a notion of content extension-
ally, in terms of what is sufficient for understanding utterances. As
Schiffer explains it, we can understand 'John said that Mary is pretty'
on Davidson's account, if we know

> (a) that ['Mary is pretty'] is true iff [Mary is pretty], (b) that that fact is entailed by a correct, finitely axiomatizable, extensional, Tarski-style truth theory for the language to which ['Mary is pretty'] belongs, and (c) that the truth theory satisfies [some appropriate] empirical constraints.

The point to-be-recognized is that average Joe does not know (b) and thus we cannot require such a theory of content in the logical form of his utterances. For surely he, and his friend average Betty, know what they are talking about without having any such truth theory. And because the revision here is on the level of logical form, whereby we come to know the meaning of the sentence, average people like Joe and Betty need to know this information explicitly, unlike, say, on the level of analysis, where such facts can be genuinely opaque to native speakers.

But the objection we should raise at this point is that Schiffer seems engaged in question-begging. The argument is that understanding 'John says that Mary is pretty' requires understanding what John said, whereas understanding 'John says that' does not. But can't Davidson just say: sure, understanding 'John says that Mary is pretty' requires one to understand 'Mary is pretty'; understanding any sequence of two sentences requires one to understand the both of them. Schiffer says that the relevant point is that we can know what John asserted without knowing the content of his utterance. Sure; but we can't know what 'John said that. Mary is pretty' means without knowing what 'Mary is pretty' means. Schiffer bases this argument on some pre-theoretical notion that distinguishes the meaning of 'John said that Mary is pretty' from the meaning of the whole 'John said that. Mary is pretty,' which is just the point at issue.

We've seen now, I think, that Davidson must retreat to utterance-kinds. But, contra Schiffer, this doesn't look like too terrible a position for him to be in. All the same, I think Davidson is wrong: the greatest appeal of his account is that it can claim that logical form is almost identical to grammatical form. This is what it means to "wear one's logical form on one's sleeve". But I think if we press the point, it becomes obvious that Davidson has to take a position where logical and grammatical forms diverge radically. 'That' is not a demonstrative

when introducing indirect discourse, and it is not a proper part of the relatum of propositional attitude ascriptions. A few observations in this direction should suffice:

For example, we don't always have a 'that,' even in cases that are clearly indirect discourse, e.g. 'John said he [John] was hungry' (if this were a direct statement, John would be saying "I am hungry"). But Davidson's account makes this statement ungrammatical: *'John said. He was hungry.' The traditional analysis of SAY has it as a transitive verb, and it doesn't seem Davidson has to posit an intransitive use. This is a plausible enough explanation: there is an unvoiced element ('that') which here occurs, the word having dropped out on account of the proximity of its referent and the rigidity of the formula in which it occurs. But that Davidson has to make such explanations is telling. It doesn't seem obvious that he can get around every objection of this sort; Consider that I can use a further demonstrative to refer to putative that-clauses, but I cannot use a second demonstrative in any other situation. We have, "I know this, that John likes Mary" though "This, that dog is big" is not the grammatical equivalent of "That dog is big." Furthermore, demonstrative 'that's—unlike, e.g., relative 'that's—can be the object of a preposition. For example, "I was unaware of that," "I was angry at that," and "I am sorry for that." But these are all infelicitous utterances: *"I was unaware of that Bush was elected," *"I was angry at that Bush was elected," and *"I was sorry for that Bush was elected," whereas these are certainly propositional attitude constructions (consider: "I was unaware that Bush was elected," or "I am sorry that Bush was elected").

More difficult for Davidson yet, 'that' doesn't seem to be a proper part of attitude ascriptions. The problem is this: adverbs and adverbial constructions may attach to the highest verb phrase in clauses of which they are a part. To give an example, 'John said that,' can have adverbial modifiers stuck on its end: 'John said that condescendingly,' or 'John said that nearly ten years ago.' Here, 'said that' is a verb phrase with 'said' as its head and 'that' as its complement, a la Davidson; 'condescendingly' and 'nearly ten years ago' hook-up with the phrase at its top (a lower node would be blocked by the position of 'that'). But this type of adverbial modification is not available to adverbials trapped after 'that's in indirect discourse. We have

(3) *John said that condescendingly he was a master chess champion.

and,

(4) John said that nearly ten years ago he was a navy seal.

which is not the same as "Nearly ten years ago, John said that he was a navy seal." However, on Davidson's account, (3) should be sensible, and (4) ambiguous. If 'that' is merely the complement of 'said,' it shouldn't block modifiers to the left attempting to attach to the verb phrase as a whole. So, in indirect discourse, 'that' as a solitary syntactic unit is not the object of the propositional attitude verb, contrary to Davidson.

Furthermore, it was pointed out by Burge that Davidson's account is suspect in that it tells native speakers that there are two independent clauses where we hear and write just one. This problem might be overlooked so long as Davidson could point to two independent clauses 'John said that,' and 'The earth moves.' But his account gets tied up when we stick on a third, dependent clause. Consider:

(5) John said that the earth moves because he wanted to sound smart.

This is a perfectly intelligible sentiment, and grammatically impeccable to boot. But not to Davidson. On his analysis:

(6) John said that. The earth moves because he wanted to sound smart.

Surely this is terrible! That's not at all what John said. Unfortunately for Davidson, there are no more moves available to him for separating the dependent clause from the second sentence. Since this dependent clause hooks-up semantically with the first independent clause, it makes sense to interpret the entire sequence 'that the earth moves' as the complement of 'said.'

In sum, I would like to remark that now it seems extraordinarily clear that Davidson's account is not one that gives a logical form admirably mirroring grammatical form. Historically, this has not been such a bad thing; but, as I shall now argue, it is damning to any theory of natural language, and doubly so for Davidson.

Russell, I think, has historically been the strongest proponent of the idea that logical form and grammatical form diverge radically, and that this should be so. Perhaps he has been misunderstood: for Russell and his contemporaries are very concerned with the forms of facts and the forms of judgments. An account of the form of a fact is an ontological account: it requires us to state in what way objects combine. This is certainly what Wittgenstein meant by 'logical form.'[1] But a sort of Quinian view of ontology precipitates a different usage of the phrase, the one that, surely, Davidson and his objectors are concerned with. Burge goes so far as to say, "Ontological positions should derive from, not dominate, the interpretation of cognitively successful discourse" (199). It is only through language that facts are expressed, and the forms of facts, on this view, just are the forms that natural language gives them. Sure, with Quine, we may change our language and thereby our ontology, but not until then are we in a position to spurn the type of logical form given us by our linguistic usage.

Even so, many philosophers believe the logical form that our language expresses is different from the surface or grammatical form it exhibits. Russell's account of definite descriptions is a tired example, and I won't go through it. But here's an interesting one: consider the old classic

(7) If God knows that I will do X, then I must do X.

It is a sentence which clearly sounds valid, but it has troubled many a scholastic, and not a few undergraduates. If the grammatical form is to-be-trusted, we are left with

(8) Knows(God, I will do X) $\rightarrow$ Necessarily(I will do X)

Well, surely God knows everything, so everything I do is of necessity and I am constrained. Modal logicians have sought to give an account of the logical form that does not result in such unpalatable consequences. Though surely it is not a move licensed by the grammar, they would like 'must' to take a wide scope over the entire conditional, as in

---

[1]The ontological character of certain uses of the phrase 'logical form' was pointed out to me by Bryan Pickel, who deserves mention.

(9) Necessarily(Knows(God, I will do X) → I will do X)

Which, under the semantics typically given for the relevant modal operators, is no unseemly conclusion after all. But even more recently a new type of account has begun to emerge, one that sanctifies grammatical form and shies away from radical reconstructions like those in (9). Suppose, for example, that we let go of the typical modal semantics and adopted an "update" semantics[2] where conditionals are true iff, updating with the antecedent, the conclusion holds. Then (8) becomes un-problematic: given that I've updated with the proposition that God knows I will do X, it is an epistemic necessity that I will do X. This doesn't entail that all my actions are constrained, only that if I know that God knows I will perform action X, then if I knew *that*, it would be necessary relevant to my knowledge base. The trick is that the problem that seemed apparent in the syntax was solved, not by re-ordering the syntax, but by complicating the semantics. This, I believe, is the first methodological decision to be made in the philosophy of language.

Davidson is clearly on the one side of the decision: he believes grammatical form should coincide as much as it can with logical form. In fact, he is committed to this position by the idea he has of how a theory of meaning for a language should look. I think, however, that every philosopher should be committed to this position. If we claim that sentences with a certain grammatical form X, have in actuality the logical form Y from which we derive their meaning by a compositional semantics, then we must hold that there is an algorithm taking us from sentences of the form X to those of the form Y. But there will be no algorithm taking us the other way 'round. Imagine how Davidson's theory of propositional attitudes might treat the French language. 'Galilei a dit que la terre bouge' might be given the logical form 'Galilei a dit cela. La terre bouge.' While consideration of the meaning of the first phrase has brought us to the second, there is no consideration of the second that can bring us to the first.

Without an algorithm from logical form to grammatical form, we may know what a sentence means (through its logical form) without

---

[2]It is from Dr. Nicholas Asher and Dr. Anthony Gillies that I have received all knowledge I have of update semantics. More than this, it is Dr. Gillies' hard line which has led me to the beliefs I have concerning logical form. They, too, deserve mention.

knowing how the familiar words contribute to its meaning. If 'the king of France is bald' has as its logical form: '(some x)(king(x, France) & (all y)(king(y, France) iff x = y) & Bald(x))' then what role is played by the familiar word 'the'? There is no proper part of the logical form that it corresponds to. Its role is relegated to hinting at the logical form, but in no other way does it contribute to the meaning of the sentence!

There is a graver worry still. If the logical form is allowed to vary significantly from the grammatical form, as when the latter can only hint at the former, what is to stop it from bringing in new ontological commitments? We must put some constraint on logical forms to square with our intuitions about the commitments of our language. What could such a constraint be? We seem to know what we are committed to existing by consideration of the meaning of our utterances. If this meaning is only arrived at through the logical form, then we should have intuitions about logical form; after all, we have intuitions about our commitments. Though it is quite clear that we have no intuitions about logical forms, at least not about those proposed above. I think most people would assent to "If God knows I will do X, then I must do X," but very few would have the intuition that this is really shortform for "It must be that if God knows I will do X, then I will do X." All our intuitions are on the side of grammatical form. This is not to say that the logical form of a fact or a judgment is here impeached—but the logical form of language is the form given by the grammar, and any theory in the philosophy of language must conform to that, or be forgotten.

So what then is the logical form of indirect discourse? I think we've been led to it quite clearly. A direct discourse statement puts forth that some state-of-affairs is the case. Singular terms have as their referents objects, and predicates represent properties of those objects. The problem we had with treating indirect contexts in this same way is Frege's problem: my believing that Hesperus is Hesperus is different from my believing that Hesperus is Phosphorus—but if the extent of our story is just that 'Hesperus' picks out an object and 'Phosphorus' that same object, and my belief is an attitude toward that object being identical to that object, then the cognitive difference goes un-captured. I shall argue though, that this is just the case.

Now, one solution on the table is to let 'Hesperus' and 'Phosphorus'

refer to different objects (e.g. senses) in indirect contexts. An obvious problem here can be seen in sentences like the following:

(10) John doesn't know that Mary is pretty, but she is.

The role of pronouns is to "pick up" the referent of an element introduced earlier in the discourse—the pronoun 'she' picks up the referent of 'Mary.' Thus if 'Mary' refers to a sense, so does 'she.' Yet we began with the observation that singular terms in direct discourse refer to objects, and it is a truly weird assertion that the sense of 'Mary' is pretty. So we must revise one of three theses: (a) that singular terms in direct discourse refer to objects, (b) that singular terms in indirect discourse refer to something other than objects (like senses), or (c) that pronouns pick up the *referents* of earlier discourse elements. (a) and (c) are rather tough to let go of. But there is worse news to be had. The whole point of senses was to allow, say, 'Hesperus' and 'Phosphorus' have separate referents in oblique contexts. But surely this is a sensible assertion:

(11) John doesn't know Hesperus is Phosphorus, but the one is the other.

Here, 'Hesperus' is not only forced to take the same referent as 'the one' (be that referent whatever it may be) but, since "the one is the other," it is forced to take the same referent as 'Phosphorus.' It seems if something must go from our theory, it's senses. So if John knows (and I'm sure he does) that Hesperus is Hesperus, it should be impossible for him not to know that Hesperus is Phosphorus, because the one is the other.

The problem, I think, is that our language is an ontological one. It's made for representing the world, not for representing our representations of the world—beliefs and desires and whatnot. Ideally we'd have two words: one for Hesperus and one for the mental representation of Hesperus. Less ideally, but still preferable, we'd have an ambiguous word that in certain contexts referred to Hesperus and in others to the representation, with those contexts being clearly determined. But as we've seen with the case of pronoun reference above, even this is not so. Here's why.

We have already noted that "Hesperus is Phosphorus" can have a different cognitive content from "Hesperus is Hesperus." But indeed

it need not. For suppose a modern astronomer, who knows beyond the shadow of any doubt that Hesperus and Phosphorus are one and the same object, wonders to himself whether Hesperus is Phosphorus. Surely to him we would attribute an interest in the law of identity. Or let's look at a clearer case. If my Latin professor were to say "Tully is a great orator," I would be correct in describing him if I said "Dr R. believes that Cicero is a great orator," whereas, reporting the same statement made by John, who is unaware of the fact that Cicero is Tully, it would be incorrect to say "John believes that Cicero is a great orator." In fact, he very well may not—he may vehemently deny it. The difference here is that while the professor has but one mental representation (encompassing two names) that corresponds to the external entity Cicero, John has two representations, each with its own name, corresponding to one and the same external entity. What we try to deliver in our belief ascription is the structure of each man's representation.

Now we are set to state clearly why there are no mental-representation terms in our language: because the sheer volume would be unfeasible. Consider: even allowing for representation types, we would need 'Cicero 1' corresponding to the person, 'Cicero 2' to representations of the professor's type, 'Cicero 3' corresponding to representations of John's type, etc. This language would be able to preserve logical inference (substitution salva veritate) but it would be monstrous, and pronominal reference to the objects of belief would have to be cut, and constructions involving them replaced with clunkier, less fluid locutions. More to the point, such a language is unlikely to have developed naturally, as it relies on a basis in the philosophy of mind—and if you thought the metaphysics of savages is bad, you should see their philosophy of mind! Hence the rather tangential use of the "object" language (no pun intended) to report representation states, like beliefs. We are still referring to the man, Cicero, when we say "John believes that Tully is a great orator," but in not allowing ourselves the logical inference to "John believes that Cicero is a great orator," we are hinting at the representational structure of John's beliefs. It is to pragmatics, and our independent understanding of belief structures and the knowledge of other minds, that we must turn, if we are to solve Frege's puzzle correctly.

Let me take this time to reiterate my position. Believing (and other

propositional attitudes) is a relation between persons and (possible) states of affairs. The logical form looks like this:

(12) Believe(John, Mary is pretty)

and it is no different from the grammatical form, and all the words have their traditional interpretations. As an analysis, we might say that (12) holds iff

(13) (some x)(x is John's sole Mary-representation & x includes the attribute "pretty") OR (some y)(y is one of John's two Mary-representations, the one indicated by the speaker of (12), & y includes the attribute "pretty") OR...

It is through pragmatic considerations that we discover which disjunct is the truth maker for (13): our knowledge about what acquaintance John has with Mary, whether she leads a double or triple life, etc.

In conclusion, Davidson's theory appealed to us because it gave us a chance to treat logical form and grammatical form as one and the same, excepting the one small point. But a closer investigation into the theory showed that logical and grammatical form diverged quite radically, sinking entirely its appeal. To find our solution we were forced to take the grammar at face value: if indirect discourse looks like direct discourse, then it is: that's all she wrote. Though this is the naive position that has been harangued against since Frege and on after, nevertheless it is the only position that can capture the data and solve the problems—and, incidentally, fit our constraint that logical and grammatical form converge.

Michael Johnson
University of Texas, Austin

Bibliography

[1] T. Burge. On Davidson's 'Saying That'. In E. LePore, editor, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Basil Blackwell, Oxford, 1986.

[2] D. Davidson. On Saying That. In *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford, 1985.

[3] S. Schiffer. Davidson on Saying That. In *Remnants of Meaning*. The MIT Press, Cambridge, 1987.

Counterexamples to Van Inwagen's Consequence Argument Reconsidered

Peter van Inwagen is an incompatibilist: he believes that free will is incompatible with determinism. According to van Inwagen, in order for the will to be free, a person must have open alternative possibilities for action. When what an agent does *actually* do and what an agent *can* do coincide, his will is not free. According to determinism, the present state of the universe is determined by the state of the universe in the past along with the laws of nature. Van Inwagen believes that if determinism is true, then all of our acts are the consequences of the laws of nature and the past. Since we have no choice about the laws of nature, and since we have no choice about the past, it seems that we have no choice about the consequences of these things (van Inwagen, 16). As a result, agents have no alternative possibilities and the will is not free.

In this paper I will consider an argument for the incompatibility of determinism and free will given by Peter van Inwagen in *An Essay on Free Will.* I will then examine a crucial step in his proof—rule (β)—more closely. In doing so, I will look at two counterexamples that seem to show the invalidity of this rule. Next, I will consider an argument, given by Michael Huemer, that allegedly makes rule (β) immune to the counterexamples. I will then argue against Huemer and offer a reason as to why his proposal must be rejected.

In arguing for the incompatibility of determinism and the freedom

of the will, van Inwagen offers an argument that relies on rules (α) and (β).

Rule (α) states:

A1. $\Box p$
A2. $\mathsf{N}p$

From A1, A2 may be inferred. $\Box$ represents logical necessity. If p is logically necessary, then no one has, or ever had, any choice about whether p. The move from 1 to 2 seems uncontroversial.

Rule (β), which is a modal argument, states:

B1. $\mathsf{N}(p \rightarrow q)$
B2. $\mathsf{N}p$
B3. $\mathsf{N}q$

From B1 and B2 one may infer B3. 'N' is a modal operator that attaches to sentences that have truth-values (van Inwagen, 93). For any p, '$\mathsf{N}p$' stands for:

p and no one has, or ever had, any choice about whether p.

According to rule (β), then: no one has, or ever had, any choice about whether if p then q, and no one has, or ever had, any choice about whether p, so, no one has, or ever had, any choice about whether q.

With these rules, van Inwagen argues as follows: Let P_o be a true proposition describing the universe at an instant in the past. Let L be all the laws of nature conjoined into a single proposition. Let P be a true proposition that describes the universe at some time after P_o. If determinism, which states $\Box((P_o \wedge L) \rightarrow P)$, is assumed to be true, it can be argued as follows:

1. $\Box((P_o \wedge L) \rightarrow P)$		*follows from the assumption that determinism is true*
2. $\Box(P_o \rightarrow (L \rightarrow P))$		*deduced from 1*
3. $\mathsf{N}(P_o \rightarrow (L \rightarrow P))$		*2; rule α*
4. $\mathsf{N}P_o$		*premise*
5. $\mathsf{N}(L \rightarrow P)$		*3,4; rule β*
6. $\mathsf{N}L$		*premise*
7. $\therefore \mathsf{N}P$		*5,6; rule β*

(van Inwagen, 94-95).

Provided that rules (α) and (β) are valid, this argument is sound and the conclusion tells us that no one has, or ever had, any control over the present state of the universe. However, while rule (α) is valid, rule (β) can be shown to be invalid.

Huemer offers a counterexample to rule (β) that is very similar to a counterexample given by Thomas J. McKay and David Johnson. I will consider the McKay and Johnson counterexample because I find it much clearer. According to McKay and Johnson's counterexample, by using rules (α) and (β) the "Principle of Agglomeration" can be derived (115). The principle states:

C1. $\mathsf{N}p$
C2. $\mathsf{N}q$
C3. $\mathsf{N}(p \wedge q)$

Here, C3 may be inferred from C1 and C2. The proof for this principle is as follows:

1. $\mathsf{N}p$	*premise*
2. $\mathsf{N}q$	*premise*
3. $\Box[p \rightarrow (q \rightarrow (p \wedge q))]$	*logical truth*
4. $\mathsf{N}[p \rightarrow (q \rightarrow (p \wedge q))]$	*3; rule α*
5. $\mathsf{N}(q \rightarrow (p \wedge q))$	*1, 4; rule β*
6. $\mathsf{N}(p \wedge q)$	*2, 5; rule β*

A counterexample to the principle of agglomeration is possible (McKay, 115). Consider the following scenario:

Some person S, who has free will, has a particle gun that shoots C-particles. Assume that a group of scientific experts have constructed the gun in such a way that it shoots particles in a predictable, but not determined, direction. When S shoots the gun, the particles are fired into a basket that is divided into a left half and a right half. If S shoots the particle gun, a particle will be fired into the basket, but it is not possible to determine whether the particle will be collected in the right half or the left half of the basket. S chooses not to fire the gun.

In this case,

$p =$ No particle lands in the left half of the basket.
$q =$ No particle lands in the right half of the basket.

In this example both premises of the principle of agglomeration are true—no one can choose in which side of the basket the particle is collected. The conclusion of the principle of agglomeration, $\mathsf{N}(p \wedge q)$, is false, though. S could have chosen to shoot the gun, in which case $(p \wedge q)$ would be false. S can choose to make $(p \wedge q)$ true or false. This is a case where both premises are true but the conclusion is false: the argument is invalid. Since a simple consequence of (α) and (β) has been derived (the Principle of Agglomeration), and from *that* a counterexample has been constructed, either (α) or (β) is invalid. Since rule (α) is clearly valid, rule (β) must be invalid.

Huemer points out that a counterexample can be made directly to rule (β) as well. For this counterexample to work, we must again assume an indeterministic world (Huemer, 532; fn.13). The world must be indeterministic because either p or q must hold, but it must be impossible to determine which will hold. Now, consider the same scenario as above and the following propositions:

$p = $ No particle lands in the left half of the basket.
$q = $ No particle lands in the right half of the basket.
$r = $ No particle lands in the basket.

In this case, S has no choice about whether p and he has no choice about whether $(p \rightarrow r)$. In order to have a choice about $(p \rightarrow r)$, S must be able to do something to bring about $\neg(p \rightarrow r)$. It can be seen that this is impossible by going through a series of equivalencies:

D1. $\neg(p \rightarrow r)$
D2. $\neg(\neg p \vee r)$
D3. $(p \wedge \neg r)$
D4. $\neg q$

D1, D2, and D3 are logically equivalent. D4 is also equivalent to D1-3, but not logically. D4 is equivalent in virtue of the story being considered—perhaps it can be called *materially equivalent*. Since S cannot guarantee $\neg q$ (which is materially equivalent in this example to $\neg(p \rightarrow r)$), S cannot do anything such that $\neg(p \rightarrow r)$ holds.

S does have a choice whether r though. If S fires the gun, a particle will land in the basket $(\neg r)$. We can see then,

1. $\mathsf{N}p$

2. $\mathsf{N}(p \to r)$, but
3. $\neg \mathsf{N} r$

Thus, we have a direct counterexample to rule (β) (Huemer, 533).

I find both of these counterexamples convincing. However, Huemer suggests that by strengthening rule (β), the counterexample can be avoided. By replacing 'p, and no one has, or ever had, any choice about whether p' with the stronger claim 'no matter what S does, p', rule (β) will be valid. Huemer gives the following definition for what I will call $\mathsf{N}p$*:

> no matter what S does, $p = p$, and for each action, A, that S can perform, if S were to perform A, it *would* still be the case that p.

Huemer then offers an example: 'No matter what I do, p' is stronger than 'p and no one has, or ever had, any choice about whether p' and $\mathsf{N}p$* can avoid the counter examples. In the example with the particle gun from above, S can shoot the gun and a particle will land in the basket. If S *did* shoot the gun, it is not the case that no particle would have landed in the left half of the basket; rather, a particle *might* have landed in the left half of the basket. This would mean that $\mathsf{N}p$* is false and by parity of reasoning $\mathsf{N}q$* is false. The scenario, then, no longer leads to true premises and a false conclusion.

While Huemer draws the correct conclusion from incorporating $\mathsf{N}p$*, there is a strong reason for rejecting its use. The stronger 'No matter what S does, p' will be unacceptable to most compatibilists on two fronts. If $\mathsf{N}p$* is true for every p whatsoever, then the only action S can perform is the one he does perform. For van Inwagen, as noted above, this is just what it means for the will *not* to be free. If this is correct, then Huemer is begging the very question at hand. $\mathsf{N}p$* will only be acceptable if it is already accepted that S can only perform the action he does perform. However, for most compatibilists, an agent's will is free when he has the ability to do what he wills to do. To make this clearer, consider the action of picking up a book. Assume it is determined that S wills to pick up a book and that this is sufficient for him to perform the act. In this case, even though S could not have done otherwise than pick up the book, since he willed to pick up the book, he picks up the book freely. So, $\mathsf{N}p$* begs the very question at

hand and, in addition, a compatibilist will argue that the will can be free even if determinism is assumed to be true.

Huemer may now argue that all I have done is point out a fundamental difference between compatibilists and incompatibilists, but I have not properly shown that he is begging the question. He might say: "When determinism is assumed, there is only one action open to S. This means that S does not have alternative possible actions and is, hence, not free. I grant that this is not what a compatibilist thinks it means for the will to be free, but this is just a fundamental point on which we differ. I am simply showing that under determinism there are no alternative possibilities. From this I conclude that the will is not free. At no point am I assuming that the will is not free. Although it appears that I have built into Np^* that S can only perform that action which he does perform, this is really a result of assuming determinism. Therefore, I am not begging the question, unless assuming determinism begs the question. You have only pointed out a fundamental disagreement between compatibilists and incompatibilists."

More can be said to Huemer though. I myself noted that Np^* has to be true for *every* p whatsoever for the above argument (the theoretical one in quotes) to be made. Suppose this is not the case and that Np^* is true only for *this* p, just this one time, but not true for other actions. If that is the case, when it comes to considering this p, S can only do this p, but he may be able to do something other than he did actually do leading up to this p. For example, agent S may be in the bathroom one morning getting ready to go to work. He brushes his teeth (action y) and then washes his face (action z). Afterward, S goes to work (action p). S could very well have brushed his teeth, shaved (action r), and then gone to work. In this scenario, S could have done some different actions leading up to p and still done p regardless of just how these actions transpired. If this is the case, it is false that the only action S can perform is the one he does perform if we are talking about action z or r, but it is true that Np^*—no matter what S does (z, $\neg z$, r, $\neg r$), he still does p.

If this is how Huemer uses Np^*, he must think that determinism is selective or fleeting. How could it be that p is going to happen no matter what while z or r may or may not happen? This could only happen if determinism were allowed to apply to some cases but not others. First, determinism is assumed to be true in van Inwagen's

argument, which means we cannot call it true sometimes and false others. Second, if something may or may not happen, then the future thereafter would not be causally determined.

Whether Huemer holds that Np^* applies to every p or just some p's[1], there is more to be said against his argument. Recall, first, that van Inwagen's argument concluded that no one has, or ever had, any control over the present state of the universe. This means that, if determinism is true, an agent can only do what he does actually do. Many compatibilists will argue that 'can' should be interpreted in a conditional way. Consider once again the action of picking up a book. Under a conditional understanding of 'can', the sentence 'I can pick up the book' means 'if I had chosen to pick up the book, I would have picked up the book'. Although this is another fundamental difference between compatibilists and incompatibilists, it is important to see why a compatibilist might favor this reading.

Consider the following scenario: This morning I got out of bed and had a bowl of cereal. This is the result of a long chain of events based on the past and the laws of nature. One of the facts from the past that directly contributed to the fact that I had a bowl of cereal this morning was the fact that I went to the store two days ago and bought cereal and milk. Had I not gone to the store in the past (if it were determined that I did not go to the store, perhaps), there would have been no cereal for me to eat this morning. It happens to be that I *did* go to the store, though. But, if the past had been different than it was, the present would be different than it is.

If this is plausible (and I believe it is), then it is clear why a compatibilist may want to accept the conditional reading of 'can'. The compatibilist is taking into consideration the fact that if the past *had* been different, then the present *would* be different as well.[2] While an incompatibilist can discard the conditional reading of 'can' as irrelevant to the problem of free will, he cannot discard the possibility that if the past had been different then the present would be different, too. To deny this is to deny the principle of determinism altogether. Thus, I believe that Huemer's Np^* denies determinism.

---

[1] "p's" is meant to mean p plural, not the possessive of p or p prime.

[2] I take it that the same is true when considering the relationship between the present and the future. If the present is different than it is, the future will be different, too.

It may be asked, "How is this any different than the case above where N*p** does not apply to every single *p*, e.g. where no matter what *S* does (*z*, ¬*z*, *r*, ¬*r*), he still does *p*?" In the case of the conditional reading of 'can', when the past is different in some way, the present will be different, too. It is possible that, ultimately, the present ends up being just as it is in reality, but this would be due to a multitude of changes in the past that ultimately negate their effects on one another. So, in this case, if the past were different, at least some instant following that time would be different, but it may then be negated. In the above example, where no matter what *S* does, he still does *p*, the present *will not* be different. The past might be different, but the present will end up being the same no matter what.

To clarify this point, and the point that N*p** denies determinism, consider Huemer's own example. He uses $2 + 2 = 4$ when considering 'no matter what *S* does, *p*'. No matter what *S* does, that mathematical truth will hold (Huemer, 538-539). This is acceptable because there is nothing *S* can do to make $2 + 2 = 4$ false. He can stand up, sit down, go to the store, or rob a bank, but $2 + 2$ will still equal 4. He could even travel back in time and blow up California, but $2 + 2$ would still equal 4. However, consider another example similar to the cereal example: being hungry. If at a certain time I am hungry and I stand up, I will still be hungry. If I walk around the room, I will still be hungry. If I rob a bank, I will still be hungry. If I eat a steak, I will *not* still be hungry. If I travel back in time and eat a steak, I will *not* be hungry. It would be odd to say that no matter what I do, I would still be hungry. The same is true for many real life cases. With the exception of necessary truths (which do not rest on causation), why assume 'no matter what *S* does, *p*'? If determinism is correctly understood, then it must be acceptable to say that if the past or the laws of nature were different in some way, the present would be different, too. N*p** is too strong to allow this. Rather than saying that the present state of the universe is determined by the past state of the universe and the laws of nature (which we have no control over but which could, theoretically, have been different), determinism would have to say something like, "the present state of the universe, which is determined by the past state of the universe and by the laws of nature, could not have been different in any way, even if the past or the laws had been different." This seems absurd.

At the end of his paper, Huemer discusses the plausibility of the Consequence Argument given by van Inwagen (now with the premise Np strengthened to Np*). In doing so, he discusses David Lewis's suggestion that "if A is any action I do not perform, if I were to perform A, some actual law of nature would be false *and* the past would be different from the actual past" (Huemer, 541).[3] Huemer then admits, "Now, it is not entirely implausible that this is true, if determinism is true" (Huemer, 542). It seems that Huemer does, at least to some extent, agree that if the past were different, the present could be different, too. Again, though, Np* makes this impossible. I do not know how Huemer can accept Np* and, at the same time, grant that what Lewis says is "not entirely implausible".

Huemer goes on to argue that despite the fact that Lewis's claim is not entirely implausible, an agent would not then be perfectly free to perform any action he likes. Although I agree with Huemer here, I still hold that Huemer has been unfair to the compatibilist and, more importantly, is in danger of falling into an inconsistent argument. Van Inwagen's argument first assumes determinism and then tries to prove that free will is incompatible with it. If Np* is used in the proof, determinism becomes implausible. Van Inwagen's Np does not run into these difficulties. For this reason, I find Huemer's move from Np to Np* unacceptable because it is incompatible with the very argument in which it is being used. I believe that van Inwagen (and most compatibilists) would agree on similar grounds.

ALEX GRZANKOWSKI
Cornell University, Ithaca

---

[3]Huemer is discussing here Lewis' 1981 article 'Are We Free to Break the Laws?' in *Theoria*, 47, pp. 113-21.

Bibliography

[1] M. Huemer. Van Inwagen's Consequence Argument. *Philosophical Review*, 109:524–544, 2000.

[2] D. Johnson and T. J. McKay. A Reconsideration of an Argument Against Compatibilism. *Philosophical Topics*, 24:113–122, 1996.

[3] Peter van Inwagen. *An Essay on Free Will*. Oxford University Press, New York, 1983.

Are Human Beings Ends in Themselves?

Much has been made of the seeming incompatibility of Kantian ethics and animal rights. Not only do several formulations of the Categorical Imperative (CI) seem to exclude animals from the sphere of moral concern, but Kant himself states that we have no direct obligations to animals at all; instead, our duties towards animals are at best indirect duties we hold towards other human beings. For example, in "Duties to Animals and Spirits" Kant writes, "[So] far as animals are concerned, we have no direct duties. Animals are not self-conscious and are there merely as the means to an end. That end is man." He reiterates this point later by writing: "Our duties towards animals are merely indirect duties towards humanity" (*ibid*). However, Kant is not always the best interpreter of his own theory, as his questionable application of the CI in *Groundwork II* demonstrates (*G* 222 [4: 421]). Moreover, he is infamous for his misogyny in *The Metaphysics of Morals* and his racism in "Observations on the Feelings of the Beautiful and the Sublime"; yet scholars do not thereby conclude that his ethical theory is sexist or racist. Therefore, we must determine whether Kant is right to exclude nonhuman animals from the moral radar. Does his view follow from a proper application of the CI, or is it simply the result of uncritical prejudice?

Animal rights advocates have argued that the Kantian assessment of animals is not only contingent, but false as well. Specifically, it is contingent in that one can easily conceive of a possible world in which

animals are rational beings[1], and it is false in that we happen to live in just such a world: not only are human beings rational animals, but substantial empirical evidence indicates that many nonhumans are as well. For example, chimpanzees, bonobos, gorillas and orangutans are able to use reason to discern cause-and-effect relationships (in a limited sense), while other animals such as cats, dogs, and pigs have intellectual capacities similar to, if not greater than, those of a normal human child. Moreover, even though not all nonhumans have a developed capacity for cognition, they are nonetheless able to set ends based on inclination and pursue the necessary means for achieving them. Therefore, we have good reason to reject the simplistic view of the animal mind that Kant describes in his books.[2]

However, it is not my present intention to side with this response, nor do I intend to develop this line of reasoning any further, as it is an empirical argument that tries to fit animals into the Kantian view, not an analysis of the view itself. Thus, for my purposes in this paper, I will presuppose that all nonhumans are nonrational—granting that this point encounters substantial empirical resistance—and argue

---

[1] Harry Frankfurt (1971) argues that a person (and thus a possessor of moral status) is an individual who can have second-order volitions, or who can desire that a particular desire "be his will" (10). Yet, even though his definition of personhood denigrates the moral status of nonhuman animals in practical terms (as they are not capable of second-order introspection), Frankfurt supports this *theoretical* point when he writes, "It is conceptually possible that members of novel or even of familiar nonhuman species should be persons; and it is also conceptually possible that some members of the human species are not persons" (6).

[2] Allen Wood (1998) encapsulates this view when he claims:

> Although nonhuman animals may not possess rational nature itself, they do possess recognizable fragments of it. They have capacities which we should value as the infrastructure, so to speak, of rational nature. Many animals have desires and they experience pleasure or pain....Many animals also have what Tom Regan calls 'preference autonomy': that is, they have preferences and the ability to initiate actions to satisfy them. Preference autonomy is not the same as the rational autonomy on which Kantian ethics is grounded, but it is a necessary condition for rational autonomy and part of its structure. (Wood 15)

Similarly, while Alan Gewirth (1978) holds that prospective purposive agency is a necessary condition for moral rights, he qualifies this claim by noting that we can create a moral heirarchy based on proportionality. That is, moral worth is present in all beings with even the most nascent capacity for reason, but only in proportion to the degree to which they have this capacity.

instead that the low moral value associated with inclination-based ends poses significant problems for the CI. Specifically, I will argue that the humanity formulation of the CI, as it is presently understood, commits Kant to the claim that human beings have indirect moral status, even though he asserts that they are ends in themselves. Then I will argue that, given that Kant denies moral worth to inclination-governed beings, the humanity formula limits us to a very confined set of duties that clashes not only with the duties we derive from the other formulations of the CI, but also with the very sensibilities on which the moral law is based. I will conclude that we cannot solve this problem unless we reframe the moral radar, and I will suggest that the only plausible means for doing so will incorporate nonrational sentient beings into the sphere of moral concern.

Given that Kant denies direct status[3] to nonhuman animals, it seems natural to begin here: What exactly does it mean to have a direct duty towards x? It means that one must treat x as an end in itself, or rather, as a being whose value commands the respect of all rational agents. In contrast, if one has an indirect duty concerning x, one must treat x in a certain way not for its own sake, but to comply with a duty one has towards y. It is in this sense that we can be said to have duties concerning nonsentient inanimate objects. For instance, we are obliged not to burn down houses, smudge paintings with our fingertips, or rip the heads off teddy bears not because these things have rights, but because they have value to people who do. Of course, that people care for these things is contingent: the same inanimate object can transit in and out of moral standing based entirely on external relational factors: for example, whether a human being values it, or whether it continues to benefit humans in general. This is why Kant considers nonsentient inanimate objects to be "relative ends" (ends whose value is dependent on our desires) rather than "objective ends" (ends in themselves).

Moreover, Kant argues that animals are relative ends in much the same way: we must treat them kindly not because they have rights that we must consider, but because they have value to people who do. In this sense, ripping the head off a kitten is no worse than ripping

---

[3]Here, I take "direct status" to be defined in terms of direct duty. That is, a person to whom we owe a direct duty can be said to have direct status. I take "indirect status" to have the same relationship with indirect duty.

the head off a teddy bear: it is wrong not because it harms the kitten, but because it harms a person who cares about the kitten.[4] However, Kant argues that harming animals is wrong for an additional reason: although it is not wrong in itself, it might habituate us to become more violent towards other human beings (as animals respond to pain in much the same way that we do). Of course, this reason is equally indirect, for even though it governs our treatment of animals, it does so only for the benefit of humans.

There seems to be something right about this. After all, it would be ridiculous to deny that animals have indirect status, at least in this general sense: if a human being cares about an animal, then harming the animal harms the human being. But having indirect status in this sense is not sufficient to render one a relative end; otherwise humans would be relative ends too. For example, if I murdered my roommate Sean, I would harm his family, his friends, and everyone else who suffered from my act. In one respect, then, Sean has indirect status: by murdering him I would breach the rights of others. But this does not mean that Sean is not an objective end, because it does not take away from the fact that *I would breach his rights too.* Everything that lives in an environmental ecosystem, let alone a social community, can be a means to other ends; but something can be a relative end if and only if it is *merely* a means to other ends. Clearly Sean fails this test, because his welfare is important to him as well as to his family, friends and community. Nonrational animals fail this test for the same reason: they are subjects of lives that are important to them. Nonsentient inanimate objects, on the other hand, have no interests at all; they are merely means towards other ends, and so they do count as relative ends.

Of course, this does not establish that animals are objective ends; to do that we need to determine whether they have *direct* status. Nevertheless, we can derive three conclusions from this: (1) people can have indirect status without thereby being relative ends; (2) sentient beings (both rational and nonrational) are distinguishable from nonsentient inanimate objects, and so the moral radar can consistently accommodate the former and not the latter; and (3) a single action

---

[4]Note that the person does not need to care *for* the kitten. For example, the rancher whose cattle die prematurely may suffer only economically, but he suffers nonetheless.

can breach the rights of multiple parties. I take the first two points to be relatively straightforward, but I believe the third requires further consideration. Specifically, I believe it raises an important question regarding the reference of moral obligations in general.

Consider the following example: when Adam promises Beth that he will drive Carl, her son, to the doctor, he is, in effect, imposing a moral duty upon himself. But to whom (or what) is Adam obliged? There are many possible responses to this question, of which I will offer four. (1) *Adam is obliged to Beth:* Adam promised Beth that he would perform the act in question, and he would be breaching a verbal contract with her if this promise were to go unfulfilled. (2) *Adam is obliged to Carl:* The duty concerns a service that Adam intends to perform for Carl, and Carl is the person who will suffer if the promise is unfulfilled. (3) *Adam is obliged to himself:* As the legislator of the duty, Adam is both its creator and its referent. Thus, even though he vocalized the promise to Beth and created it concerning Carl, it refers to no one but himself. (4) *Adam is obliged to his community:* The referent of his duty is not a particular person, but rather the community as a whole. For even though promise-keeping might yield negative results in particular cases, its continued occurrence will lead to maximum benefit in the long run.

Each of these responses is intuitively plausible, and they all find strong support in the *Groundwork*. Specifically, (1) follows from the argument that a maxim of false promising fails the contradiction in conception test (*G* 223 [4: 422]); (2) follows from the argument that we must treat humanity not merely as a means but as an end in itself (229 [4: 428]); (3) follows from the argument that autonomy of the will is the ground of moral obligation (231 [4: 431]); and (4) follows from the argument that we must judge the maxims of our actions from the perspective of a kingdom of ends (233 [4: 433]). Consequently, if we are to maintain coherence in the Kantian scheme, we must conclude that Adam can hold all four obligations at once. The question we must address, then, is not *which* duty Adam holds, but rather whether or not the four duties are compatible with one another, and whether each is a distinct duty or a distinct description of the same duty.

I believe that an ambiguity in the CI prevents us from finding a clear answer to this question. In order to demonstrate this point, allow me to alter the example as follows: Adam promises Beth that he

will drive Dumbledore, her dog, to the vet. In this case, the same reason that obliges Adam to Carl in response (2) obliges him to Dumbledore as well: the duty concerns a service that Adam intends to perform for Dumbledore, and Dumbledore is the individual who will suffer if the promise is unfulfilled. So, if Adam has a direct moral obligation to Carl, then he must also have a direct moral obligation to Dumbledore—unless a relevant distinction separates the two cases. In other words, if we are to maintain the Kantian distinction between humans and nonhumans, the source of duty for Adam must be something other than the fact that Carl is the most direct beneficiary of his action, as this condition applies equally to Dumbledore.

In response to this challenge, Kantians have defended the demarcation between humans and nonhumans by citing the humanity formulation of the CI, which claims that humanity is an end to which we hold direct moral duties. In *Groundwork II*, Kant writes:

> *Rational nature exists as an end in itself.* This is the way in which a human being necessarily conceives his own existence, and it is therefore a *subjective* principle of human actions. But it is also the way in which every other rational being conceives his existence, on the same rational ground which holds also for me; hence it is at the same time an objective principle from which, since it is a supreme practical ground, it must be possible to derive all laws of the will. The practical imperative will therefore be the following: *Act in such a way that you treat humanity, whether in your own person or in any other person, always at the same time as an end, never merely as a means.* (*G* 229-230 [4: 429])

It is in this passage that Kant distinguishes relative ends from objective ends, arguing that while the former "provide no universal principles, no principles valid and necessary for all rational beings and also for every volition," the latter are "valid for all rational beings" (*G* 228 [427]). Furthermore, Kant writes, "[As] rational beings, [human beings] must always at the same time be valued as ends"; and, "[The] principle 'So act in relation to every rational being (both yourself and others) that this being may at the same time count in your maxim as an end in itself' is...basically the same as the principle, 'Act on a maxim which

at the same time embodies in itself its own universal validity for every rational being'" (*G* 230 [4: 430], 238 [4: 438]).

It is this distinction, Kant believes, that allows us to remove non-rational animals from the sphere of moral concern: Adam is obliged to Carl (and not Dumbledore) *not* because of the particular circumstances surrounding his promise, but because Carl is a member of the human race—a class of beings[5] that, due to their rational nature, are objective ends, or ends in themselves. In contrast, Dumbledore is part of a species that has mere price instead of dignity[6], because its members are capable of responding only to inclination, and dignity stems only from the presence of humanity.[7] Thus, even though Dumbledore is a sentient being capable of pleasure, pain and happiness, he is not an objective end: his value is dependent on the desires of humans.

It is important to emphasize that according to Kant, our worth as ends is predicated on the fact that we possess a capacity for reason. As Thomas Hill notes in his article 'Humanity as an End in Itself' (1980), Kant repeatedly uses the phrase "humanity in a person" in *The Metaphysics of Morals*, suggesting that humanity is not a description of a particular species, but rather a property found in all people— whether or not they happen to be biologically human (Hill 85). Kant also supports this interpretation when he introduces the humanity formula in *Groundwork II*, writing, "*Rational nature* exists as an end in itself," and, "The practical imperative will therefore be the following: Act in such a way that you treat humanity, whether *in* your own

---

[5]I take 'class of beings' and 'species' nominalistically, i.e., shorthand for all the members of the group, rather than the group as something over and beyond its members.

[6]*G* 235 (4: 434).

[7]Barbara Herman (1984) represents this view, arguing that "objects and animals cannot respond to need as such, nor can they take my ends as their own," and, "animals are not, strictly speaking, capable of providing help, although they may of course do things that are helpful to us" (Herman 588, 593). As a result, Herman concludes (based on Kantian grounds, no less) that the positive duty of beneficence does not apply to nonrational beings.

Other philosophers who have supported this position include Henry J. McCloskey (1979), who argues that either actual or potential moral agency is a necessary condition for moral worth, and Carl Wellman (1995), who believes that the freedom to control oneself in a possible confrontation is a requirement for moral worth. Thus, neither McCloskey nor Wellman believes that nonhuman animals have moral worth.

person or *in* any other person, always at the same time as an end, never merely as a means" (*G* 229-230 [4: 429], emphasis mine). This passage implies that our moral duties refer not to ourselves, nor to those around us, nor even to the communities in which we live; rather, they refer to nothing other than *reason itself*—the rational nature that permeates us and allows us to act out of reverence for the moral law. And since humanity, which affords us dignity instead of mere price, is coextensive with, if not identical to, rational nature, *it is the rational nature in people that possesses dignity, not the people themselves.*

Further passages in the *Groundwork* support this analysis. For example, in *Groundwork I*, Kant argues:

> Now if an action done out of duty is supposed to exclude totally the influence of inclination, and, along with inclination, every object of volition, then nothing remains that could determine the will except objectively the law and subjectively pure respect for this practical law. (*G* 202 [4: 400])

Moreover, he claims:

> That preeminent good which we call "moral" consists therefore in nothing but the idea of the law in itself, which certainly is present only in a rational being—so far as that idea, and not an expected result, is the determining ground of the will. (*G* 203 [4: 402])

Kant believes that this argument grounds the claim that human beings are ends in themselves: we have moral worth, Kant asserts, because we are imbued with rational nature. In this sense, the humanity formula helps us to understand how Adam can legislate four seemingly distinct duties by means of a single promise. In promising Beth that he will drive Carl to the doctor, he assumes a duty analogous to a four-place predicate, or a function involving four unbound variables: Adam is obliged (equally) to Beth, Carl, himself, and his community[8] because all four are connected to rational nature, to which Adam has a direct moral duty. Moreover, Kantians might argue, the humanity formula helps us to understand why Adam has a direct obligation to

---

[8]Given the definition of "community" in the Kingdom of Ends formula.

Carl but not to Dumbledore. Namely, in the latter case, Adam is obligated only to Beth, himself, and his community because Dumbledore is nonrational and thus outside the sphere of moral concern. However, I am not sure that this distinction holds—not because I believe that Dumbledore has direct moral status, but because I believe that Carl might not.

Kant introduces the humanity formula in part because he wants to avoid the speciesist implication that humanity is limited to human beings. As Hill discusses in his edition of the *Groundwork*, Kant intends the humanity formula to expand the moral radar such that it includes any being with a capacity for reason, such as angels or rational aliens (269). However, by preserving the possibility that humanity is multiply realizable, Kant imposes serious constraints on the direct duties we can derive from the humanity formula. In order to understand why, we must consider the *Groundwork III*, where Kant discusses our rational nature:

> [A rational being] has two perspectives from which he can consider himself and from which he can acknowledge the laws governing the use of his powers and consequently governing all his actions. He can consider himself first so far as he belongs to the world of sense, under laws of nature (heteronomy); and secondly—so far as he belongs to the intelligible world—under laws that are not empirical but, being independent of nature, are founded on reason alone. (*G* 252 [4: 452])

Kant distinguishes between autonomy and heteronomy to support the argument that we must imagine ourselves as transcendentally free, for otherwise we would not be subject to the moral law. However, he grants that we must also imagine ourselves as part of the causal order, and that the perspective from which we are free is no more compelling than that from which we are not. It is this point that presents a problem for the notion that our moral worth is predicated on our humanity, for if it is true that we must imagine ourselves as both autonomous and heteronomous, and if it is also true that moral duty refers directly (and exclusively) to rational nature, then it follows that we have direct moral duties only to our *autonomous* selves. That is, whereas we have a direct moral duty to cultivate and preserve our

rational nature, we have, at best, an indirect moral duty to care for our physical bodies. After all, they reside in the world of sense, and so they, along with all the ends they pursue, are subject only to the laws of physical causation. Thus, they have mere price, not dignity.

Accordingly, all our negative moral duties must fall under two main categories: the duty to avoid maxims that insult or frustrate the process of reason, and the duty to avoid maxims that hinder the physical processes required for rational behavior. For example, it is always wrong to tell a lie or make a false promise, because such behavior misrepresents the world to other rational beings and thereby insults the dignity of rational nature. Kant argues for this conclusion through the false promising example in *Groundwork II*, as well as through arguments elsewhere condemning lying, mockery, and servility. As Hill notes, "Kant is unusual, at least compared to moral philosophers today, in stressing the moral importance of attitude and gesture aside from their consequences. Mockery is opposed, whether or not it is effective for the purpose of reform or deterrent, because it reflects a disrespectful attitude toward the humanity of others" (Hill 97). Additionally, under this view, it is always wrong to consume drugs and alcohol, because they seem to negatively impact our capacity to perform higher cognitive functions. Kant argues for this conclusion in many passages as well, claiming, for example, that one should never consume opiates or alcohol because they cause one to be temporarily irrational, with a weakened "capacity to use his powers purposively" (*MM* 180 [427]).

On the other hand, we have no negative duties to avoid any act that resides exclusively in the domain of inclination. For example, I have no duty to stop myself from chopping off your arms and legs, because such an act would not hinder your capacity for reason in the least. In fact, the amputation of certain body parts would lead to a *decline* of the temptations associated with inclination, thereby allowing reason to hold a firmer grasp on your motives.[9] Granted, one could argue that our heteronomy has moral value because of its connection to our autonomy; but why should an inclination have *direct* moral value simply because it resides in a body that also happens to house a capacity for reason? This would be tantamount to arguing that

---

[9]Consider, for example, the forcible castration of pathological sex offenders.

Dumbledore has direct moral value because he lives in a house that also accommodates Beth. In other words, the connection between our autonomy and our heteronomy is every bit as indirect as the connection between rational agents, nonrational sentient beings, and nonsentient inanimate objects. Consequently, the humanity formula reduces all duties not directly concerned with reason to indirect moral status.[10] It is in this respect that the humanity formula clashes with the other formulations of the CI, as well as with our moral sensibilities; for it commits us to an untenable demarcation of direct and indirect moral duties. For example, we have a direct duty not to anesthetize patients prior to surgery, yet we have a mere indirect duty not to rape.

This last point, of course, is inconsistent with the universal law formulation of the CI, from which we derive a direct moral duty not to rape. Since rape is an interpersonal act predicated on unwilling submission, it is impossible for one to rape another without failing the contradiction in conception test; for by universalizing rape, one necessarily wills to be an *unwilling* participant in the act. Thus, one cannot coherently imagine rape as a universal law, and so the maxim fails under the CI.[11] If, however, we interpret the humanity formula

---

[10]This argument is similar to that which Tom Regan (1983) develops against utilitarianism. Regan argues that utilitarianism provides inadequate grounding for ethical obligation because it reduces the status of human beings to that of mere receptacles of pleasure and pain. That is, according to utilitarianism, we are important not inherently, but because we add or detract a given amount of value from the aggregate good. Similarly, I am arguing here that the humanity formula reduces human beings to mere receptacles of rational nature; we are important not as ends in ourselves, but because we are vessels for reason. Yet, the relegation of human beings to indirect moral status conflicts with the rest of the Kantian scheme. As Regan argues, we should have moral rights because we are the subjects of lives that are important to us, not because our bodies are vehicles for that which *does* have moral worth. Thus, the moral radar should encompass not only normal human beings, but also exceptional human beings, nonrational animals, and all others who satisfy this condition.

[11]My implementation of the universal law formula in this section is based on the interpretation Richard Galvin (1991) offers regarding the contradiction in conception test. Supporting a position previously argued for by Marcus Singer in *Generalization in Ethics* (New York, 1961), Galvin criticizes three general views of the contradiction in conception test, including the "Causal Law Versions," the "Teleological Law Versions," and the "Inconsistency of Intentions Views." He then defends as an alternative the "Strict Logical Impossibility View," which derives the immorality of an act from the form of its maxim, not from the synthetic circumstances which surround it. Finally, Galvin illustrates the superiority of the "SLI"

as denying direct moral worth to our heteronomy, then this argument loses its force. After all, the desire for rape (as well as the aversion to it) is a physical drive that has no direct impact on our capacity for reason. Of course, one could argue that in the case of rape, the pain and trauma incurred renders the victim less capable of acting rationally, but this consequence is contingent and not unequivocally true. For example, people often claim that traumatic events serve as *catalysts* for rational behavior, helping them to reprioritize their lives and focus on what is important. Moreover, some philosophers even claim that pain and suffering is an important channel for rational development. Aristotle, for example, argued that trauma is a cathartic experience that allows us to overcome emotive impulses in the future, while Nietzsche argued that masters of morality flourish in life only by overcoming severe physical impediments.

Therefore, if we are to deem rape immoral with the humanity formula (and thus maintain coherence within the CI), we need to do so based on grounds other than that it has the potential to hinder our rational agency. This demonstrates that while rational nature may be the *source* of the duty not to rape, it cannot be the *referent*. The referent must be an inclination-based end to not be raped, or perhaps more generally, not to be abused and exploited. But if this is the case, then we cannot maintain the duty not to rape unless we extend its scope to nonrational sentient beings as well, since they also desire not to be abused and exploited. Therefore, if rape is immoral, then rational nature cannot be a necessary feature of the ends we endorse (e.g. the end to not be raped); rather, it is the tool that determines whether we should endorse them in the first place. Then, once an end meets the requirements of the moral law, it commands our respect wherever it may be found—be it in a rational agent or in a nonrational sentient being. Consequently, all permissible inclination-based ends are in the same moral boat: how we treat one determines how we treat the rest.

We are thus faced with a dilemma: either we uphold rational nature as both the source and referent of our duties, thereby accepting the permissibility of all acts that do not directly hinder it, or we uphold rational nature as the source of our duties but not as the referent,

---

interpretation by applying it to the cases of false promising, lying, rape, slavery, stealing.

thereby accepting that it need not be present in the objects of our duties. Our decision regarding this dilemma will be based on whether or not we choose to reconsider the moral status of inclination. On one hand, if we decide that all inclination-governed beings have the same low status ascribed to Dumbledore, then we must extend this low status to the perspective from which we view ourselves as heteronomous. On the other hand, if we revaluate heteronomy such that we provide a binding proscription of acts such as rape, then we must apply this revaluation to all heteronomous ends, whether or not they happen to reside in a creature with a capacity for reason. We must, in other words, find Kantian grounds for distinguishing moral agents from moral patients.

I believe that the second response is not only more plausible, but also more consistent with the rest of the Kantian scheme. Most notably, it coheres with the argument in *The Metaphysics of Morals* that we have a moral obligation to pursue the happiness of others. Kant writes, "When it comes to my promoting happiness as an end that is also a duty, this must therefore be the happiness of *other* human beings, whose (permitted) *end I thus make my own end as well*" (*MM* 151 [6:388]).[12] Kant even specifies that the type of happiness in question is "natural happiness (which consists in satisfaction with what nature bestows, and so with what one *enjoys* as a gift from without)" (*ibid*). Thus, according to Kant, we have a moral duty to promote the permissible inclinations of human beings, or the permissible ends of our heteronomous selves. And given that one need not be rational in order to have heteronomous ends, the following question is raised: is the duty to pursue the permissible ends of others limited to the community of moral agents, or does it extend to all heteronomous creatures?

The answer to this question hinges on what is required for an end to be "permissible" in the relevant sense. Several sections in the *Groundwork* suggest that an end can be permissible even if its possessor is not a moral agent. For example, Kant states in the *Groundwork II* that for an end to be permissible, it must be pursued not out of reverence for the CI, but only in accordance with it: "an action that is com-

---

[12]This duty does not extend to our own ends simply because we are naturally inclined to pursue them, and Kant believes that the concept "duty" implies constraint.

patible with the autonomy of the will is *permitted*; one that does not harmonize with it is *forbidden*" (*G* 240 [4:439]). In other words, an end is permissible if and only if one can measure it against the moral law without exposing a fatal contradiction, either in conception or in will. This requirement, however, does not imply that the being whose end is in question must *engage* in such measurement, or that the end must be pursued out of respect for, or even in recognition of, the moral law. One does not render an end impermissible, after all, merely by pursuing it out of inclination. For if this were the case, then we would have a moral duty to deny ourselves and others of *all* ends based on inclination—up to and including happiness itself. In other words, we would have to ensure that everyone in the kingdom of ends lives the life of an ascetic. And clearly, when Kant argues that we have a duty to pursue the permissible ends of others, he does not mean that we have a duty to prevent others from being happy. Rather, he means that when the ends of others are consistent with the requirements of reason, we have a duty to pursue them.[13]

Therefore, Kantian ethics allows for nonrational sentient beings to have permissible ends in the relevant sense. All an end needs to be permissible is this: the end, along with all the maxims of its necessary means, must be in compliance with the CI. That its possessor acts out of reverence for the moral law, on the other hand, is irrelevant (at least within the confines of our present discussion): it is important only in determining the moral standing of the possessor. Hence, Kantian ethics not only allows but demands that the domain of ends extend beyond the community of moral agents: moral agents have a duty to promote the permissible ends of *all* creatures driven by inclination and capable of pleasure, pain, and happiness.[14]

The claim that moral agency is not necessary for moral worth has roots in contemporary Kant scholarship as well, especially regarding the purpose of the CI. For example, Barbara Herman (1984) writes,

---

[13]I take the duty to pursue the permissible ends of others to be a positive duty of imperfect obligation, similar to the duty of beneficence, since one could possibly pursue every permissible end in the world. However, I take the duty to respect the permissible ends of others to be a negative duty of perfect obligation, similar to the duty of nonmaleficence, such that one can never breach it under any circumstances.

[14]Note that this argument does not preclude the possibility of there being degrees of patienthood.

"I view the CI procedure as being designed to draw our attention to those features of our condition—as rational agents in this world and as members of a community of persons—that serve as the conditions of our willings" (584). Here, it is significant that Herman sets "rational agents in this world" apart from "members of a community of persons," as it suggests that the two categories are not necessarily identical. And if Kantianism allows for an individual to be a "member of a community of persons" without also being a "rational agent," then we might have Kantian grounds for distinguishing moral agents from moral patients—which is necessary if we are to assert that nonrational beings can be objective ends.

Herman hints at this possibility when she argues that angels do not have a positive duty of beneficence towards human beings because they "are not vulnerable and dependent," and so "the argument for beneficence could not require them to reject a maxim of nonbeneficence" (590). In other words, since angels do not require help from others, they may choose not to lend aid to others without fear of contradiction. In contrast, Herman argues, since human beings are both autonomous and heteronomous, they must reject the maxim of nonbeneficence in order to pass the contradiction of will test. She writes, "The CI procedure is to show that, for any of us [human beings], the availability of the help of others is not something it can be rational to forgo," and that, "[The] argument that defeats the maxim of nonbeneficence leads, positively, to a duty of mutual aid" (584, 592).

Interestingly, this passage implies that if angels *were* "vulnerable and dependent," they would have a positive duty of beneficence towards human beings. For even though they (presumably) have a superior capacity for reason, they would still be obliged to us because of our shared need for help and our causal efficacy towards one another. In order to spell out this point, let me emphasize what it is not asserting. First, it is not suggesting that human ends would translate to angel duties only if the angels valued them too. Even if the angels had no respiratory systems, for example, they would still not be permitted to rid the universe of oxygen—just as I would not be permitted to break your glasses even if I had perfect vision. Simply put, the angels would be obliged to help us not because *they* valued our ends, but because *we* did. Second, this point is not implying that the angels would hold duties to us only if they depended on us. Even if they were not in the

position to benefit from us at all, they would still not be permitted to ignore our needs, or to exploit us for personal gain. This is because for them, the grounding of the duty of beneficence would not be a contractualist desire for reciprocity; it would be a logical consequence of the value they place on help in general. In other words, they would be obliged to help us because they would be dependent on *others*, not because they would be dependent on *us*. As Herman argues:

> [It] is the fact of our dependency—that we are, equally, dependent (again: not that we are equally dependent)—that is the ground of the duty to help. I may not be indifferent to others not because I would thereby risk the loss of needed help (this is not a duty of fairness or reciprocity) but because I cannot escape our shared condition of dependency. (592)

This argument carries with it important implications regarding nonrational beings, because the relationship between "vulnerable and dependent" angels and humans is analogous to that between humans and nonhuman animals. After all, even though human beings have a superior capacity for reason compared to animals, we are nevertheless bound to them by virtue of our shared heteronomous needs—"those features of our condition...that serve as the conditions of our willing." Thus, even though animals might not be in the position to help us, they still have ends to which we hold moral duties. In this respect, Herman provides us with a Kantian precedent, or at least the beginnings of a Kantian rationale, for the claim that a being can have moral rights without having moral duties, namely, while both autonomy and heteronomy are necessary for one to be a moral agent[15], heteronomy

---

[15]This point mirrors passages in *Groundwork II* that claim that only beings that are both autonomous and heteronomous are subject to duty, since the purely heteronomous do not have transcendental freedom, and the purely autonomous are unmoved by the influence of inclination. For example, Kant writes, "[Autonomy] of the will is unavoidably bound up with [morality] or rather is its foundation" (245 [4: 445]), as well as:

> A will whose maxims necessarily agree with the laws of autonomy is a *holy*, absolutely good will. The dependence of a will not absolutely good on the principle of autonomy (that is, moral necessitation) is *obligation*. Obligation can thus not apply to a holy being. The objective necessity of an action out of obligation is called *duty*.

alone is sufficient for one to be a moral patient, or to have ends that a moral agent must consider.[16]

This interpretation of the humanity formula also provides a solution to the Kantian problem of "exceptional" human beings such as infants, the elderly, and the severely mentally retarded. As Hill notes, "a serious worry" about the humanity formula "is that it places a comparatively higher value on rational capacity, development, control, and honor than most morally conscientious and reasonable people are prepared to grant" (Hill 98). Indeed, Kantians who wish to ascribe moral status to exceptional human beings are forced to confront a serious overemphasis on rationality in the text, and any solution predicated on a version of the humanity formula that preserves moral worth only for the rational is destined to fail. For example, Herman tries to resolve the issue by arguing that infants have the *potential* to develop a capacity for rationality, claiming, "We might regard an infant as one whose present inability to help will be overcome in the passage of time" (Herman 593). However, not only is this claim not true in all cases—and not applicable to the severely mentally retarded—but we have good reason to reject the argument that we can attribute rights to beings based on rights they may one day have. After all, we do not allow law students to convict criminals in the courtroom, nor do we permit medical interns to perform open-heart surgery. As Peter Singer puts it, a prince, by virtue of being a potential king (and having a high probability of actualizing this potential), does not have the rights of a king. Given this, why should we allow nonrational infants the same rights as rational beings—especially if we grant rationality as a prerequisite for rights in the first place?

(240 [4: 440]).

[16]For further (non-Kantian) arguments regarding this point, refer to Mac-Cormick (1976) and Feinberg (1980). MacCormick argues that to have a right is to have an interest that should be protected by the duties of others, and Feinberg contends that to have a right is to be a claimant, or to have interests on behalf of which others can speak. This interest theory of rights is compatible not only with animal rights, but also with the rights of human infants and groups of people. I should note, however, that Feinberg qualifies this claim by arguing that *species* of animals cannot have any rights—not even the right to survive—because they do not have collective interests. This is an important point, especially when our discussion enters the realm of endangered species preservation. For my purposes in this paper, though, I will limit our discussion to the rights of *individual* animals.

Laurence Thomas (2003) recognizes this problem, writing:

> Could individuals [who have severe mental retardation]...see
> themselves as Legislators of universal moral law or as mem-
> bers of the Kingdom of Ends? Would it be reasonable to
> expect them to see themselves in this way? Could they
> be expected to have self-respect on the basis of reasoning?
> Just what does it mean for others to have Kantian respect
> for SMR persons? (Thomas 9-10).

However, Thomas offers a solution that also preserves the logocentric
nature of the humanity formula. He argues: "[In the case of SMR
persons] the mentally deficient is a person for whom something went
terribly wrong. There is something that he ought to have (or should
have had) but lacks, which is very different from merely lacking some-
thing" (22). Thus, Thomas sets apart rational and nonrational sen-
tient beings by referencing *teleology* as the relevant distinction: the
severely mentally retarded might have the same intellectual capacities
as nonhuman animals, but they were *meant* to be rational, and so
we should treat them as if they are. However, this argument is even
weaker than the previous one, for not only does Thomas neglect to
explain the nature of this teleology, but he also refuses to explain why
it is morally relevant. After all, if I were to inject a rock with a "ra-
tionality serum," would it then have direct moral status *even if* my
experiment failed and it never gained the capacity for rationality?[17]

---

[17]This argument owes a significant debt to Michael Tooley, who develops a sim-
ilar line of reasoning in his article, "In Defense of Abortion and Infanticide." He
writes:

> Suppose at some future time a chemical were to be discovered which
> when injected into the brain of a kitten would cause the kitten to
> develop into a cat possessing a brain of the sort possessed by humans,
> and consequently into a cat having all the psychological capabilities
> characteristic of normal adult humans. Such cats would be able to
> think, to use language, and so on. Now it would surely be morally
> indefensible in such a situation to hold that it is seriously wrong
> to kill an adult member of the species Homo sapiens without also
> holding that it is wrong to kill any cat that has undergone such a
> process of development: there would be no difference. (205)

With this foundation in place, Tooley introduces a series of thought experiments
in support of the thesis that killing an injected kitten would be morally equiva-
lent to killing a normal kitten; and so potential agency—produced "naturally" or

Surely not. It is not a dormant or stunted capacity for rationality that gives exceptional humans value; it is the fact that they are sentient creatures with the capacity for pain, pleasure, and happiness.

In this paper, I have argued that the humanity formula of the CI commits us to accepting that our heteronomy, and all the ends it pursues, have indirect moral status. Therefore, the duties we hold towards one another based on maxims unrelated to reason are no more binding than the ones we hold towards nonrational sentient beings. As a result, either we must accept that our heteronomous selves have the same low status that we attribute to nonrational beings, or we must reconsider the idea that heteronomy carries with it a morally degrading connotation. In response to this dilemma, I have argued that the second option is not only the more plausible of the two, but also the most consistent with the primary text and the secondary literature, as well as the best means for resolving the Kantian issue of exceptional human beings.

If all this is true, then, what becomes of human beings and nonhuman animals? Who has direct status and who does not? The answer, I believe, is that the moral radar must encompass more than rational nature itself. We have a duty to pursue all permissible ends, including those that are not directly connected to rational nature. Does this mean that nonrational sentient beings are objective ends? We might be tempted to say no, because as moral agents, we act according to the demands of reason, and we do so out of pure reverence for the moral law. However, we need to recognize that even though rational nature is the source of moral obligation, it is not its referent: through our acts of willing, we endorse and pledge to preserve all permissible inclination-based ends, whether we find them in other moral agents or in nonrational sentient beings. This obligation is a logical consequence of the value we place on heteronomy in general, and it is vital to our experience as dependent beings in a world of limited resources

---

otherwise—is morally irrelevant. To this effect he argues: "It perhaps needs to be emphasized here that the moral symmetry principle does not imply that neither action is morally wrong. Perhaps both actions are wrong, even seriously so. The moral symmetry principle implies only that if they are wrong, they are so to precisely the same degree" (*ibid*). Thus, even though Tooley develops this argument as a means for addressing abortion and euthanasia, it is relevant for this discussion as well, as he grounds it on the premise that teleology is not an adequate surrogate for direct status.

and unlimited dangers. So, are nonrational sentient beings ends in themselves? To the extent that we are, yes.

Jeff Sebo
Texas Christian University, Fort Worth

Bibliography

[1] J. Feinberg. *Rights, Justice, and the Bounds of Liberty*. Princeton University Press, Princeton, 1980.

[2] H. Frankfurt. Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, 68:5–20, 1971.

[3] R. Galvin. Ethical Formalism: The Contradiction in Conception Test. *History of Philosophy Quarterly*, 8:387–408, 1991.

[4] A. Gewirth. *Reason and Morality*. University of Chicago Press, Chicago, 1978.

[5] B. Herman. Mutual Aid and Respect for Persons. *Ethics*, 94:577–602, 1984.

[6] T. Hill. Humanity as an End in Itself. *Ethics*, 91:84–89, 1980.

[7] I. Kant. Duties to Animals and Spirits. In *Lectures on Ethics*, trans. by L. Infield. Harper and Row, New York, 1963.

[8] I. Kant. *The Metaphysics of Morals*. trans. by M. Gregor. Cambridge University Press, Cambridge, 1996.

[9] I. Kant. *Groundwork for the Metaphysics of Morals*. trans. by A. Zweig. Oxford University Press, Oxford, 2002.

[10] N. MacCormick. Children's Rights: A Test-Case for Theories of Right. *Archiv fuer Rechts- und Sozialphilosophie*, 62:305–317, 1976.

[11] H. J. McCloskey. Moral Rights and Animals. *Inquiry*, 22:23–54, 1979.

[12] T. Regan. *The Case for Animal Rights*. University of California Press, Berkeley, 1983.

[13] L. Thomas. Kant on Equality. *Cognitive Sciences Eprint Archives*, 3176, October 4, 2003.

[14] M. Tooley. In Defense of Abortion and Infanticide. In J. Feinberg, editor, *The Problem of Abortion*. Wadsworth, Belmont, 1984.

[15] C. Wellman. *Real Rights*. Oxford University Press, 1995.

[16] A. Wood. Kant on Duties Regarding Nonrational Nature. *Supplement to the Proceedings of the Aristotelian Society*, 72:189–210, 1998.

The Philosophy of Mathematics in the *Tractatus*

Numerals occur in two kinds of sentence. One kind, (α), are of the form 'There are $\underline{n}$ F's in C': '$\underline{n}$' is a numeral, 'F' a count noun, and 'C' a container phrase as I shall call it. To determine the truth-value of an α, for example, 'There are 4 red circles on the wall', one checks to see if exactly 4 things satisfying the compound predicate, being red and circle, also satisfy being on the wall. The truth of an α depends on "how things are in the world." α's are contingently true or false. The other kind, (β), are of the form '$\underline{m} = \underline{n}$': '$\underline{m}$' and '$\underline{n}$' are two purely numerical expressions. To determine the truth-value of a β, for example, '$7 + 5 = 12$', one does not check the world. β's are necessarily true or false. The contingency of α's and the necessity of β's implies that they differ epistemologically: the former are *a posteriori*, the latter *a priori*. That is, the truth maker for a β is a feature of its truth bearer, whereas the truth maker for an α is different from and external to its truth bearer.

A philosophical account of α and β sentences needs to answer three questions. (i) What is the ground for applying an $\underline{n}$ to a collection? An explanation of this ground will answer the traditional question: What is a number? (ii) What is used for determining the truth-value of a β? (iii) What does the use of a numeral in an α have in common with the use of that numeral in a β? In this paper, I will explain how Wittgenstein addresses the first of these questions. According to

my reading of the *Tractatus*, the numerals in the α's do not apply
to a collection on the basis of that collection's having a "property."
In developing my reading, I will compare the position I attribute to
Wittgenstein with one that I attribute to Frege.

A treatment of the α's needs to present an account of their truth
makers. In order to do this, I will introduce a device meant to bring out
ideas about the world suggested by, but different from, Wittgenstein's
picture theory. Imagine a sheet of paper on which is written a list of
strings of marks. This list goes proxy for the world. Let the strings be
composed of three kinds of marks. Marks of the first kind are instances
of 'x_i'; those of the second kind instances of 'F_i'; and those of the third
kind instances of 'R_i'. Each string is a combination of these marks.
The strings on the list are either of form '$F_i x_j$' or of form '$R_i x_{j1}...x_{jn}$'.

I need now to remark on how I will let the list go proxy for the
world. Each mark on the list stands for an "entity" in the world. An
'x_i' stands for a particular; an 'F_i' for a property; and an 'R_i' for a
relation. I use these category terms merely to facilitate the discussion
and not to pretend to have solved any philosophical problem. The list
is composed of strings of marks. These strings are not to be thought
of as true or false. They are meant to "represent" how it is in the
world. The strings stand for combinations of the entities stood for by
the marks. These combinations may be called facts, though there is
no need for a category term.

I said the list goes proxy for the world. That is, the list "represents"
the world. It is complete: it presents all the facts in the world and
only those facts. There is nothing else on the list, and according to
Wittgenstein, there is nothing else in the world. Further, there is
nothing to be added to the list. It will be the device that I will use
to discuss the truth makers for ordinary language sentences. Most
importantly, one must not suppose that there are strings about the
list on the list, such as that the list is complete.[1]

I am going to use the list to explain how philosophers have used
make-true talk, in particular, of course, for the α sentences. The reason
for introducing the completeness of the list is to give sense to viewing
the world *sub specie aeternitatis*. Finally, in discussing the truth or

---

[1] Because this list is primarily designed to address epistemological issues related
to the sentences, it does not take up the discussion of possibility—the notion that
is central to the picture theory of language.

falsity of an α, I will treat the list as if it were the world. The world, so to speak, is on the page.

Let me illustrate how I use the list to explicate truth-maker talk about sentences in the ordinary language. Consider the sentences: 'John is a chair in the room'; 'Harry is a chair in the room'; 'Eric is a chair in the room'; and 'Karl is a chair in the room'. If these sentences are true, the names stand for the four marks: 'x_1', 'x_2', 'x_3', and 'x_4' respectively;[2] 'the room' for 'x_5'; 'chair' for 'F_1'; and 'in the' for 'R_1'. Further, the following strings must be on the list: 'x_1F_1'; 'x_2F_1'; 'x_3F_1'; 'x_4F_1'; '$x_1R_1x_5$'; '$x_2R_1x_5$'; '$x_3R_1x_5$'; and '$x_4R_1x_5$'. If any of these strings is not on the list, then one of the corresponding sentences is false. I, of course, make the simplifying assumption that these English words are correlated to marks on the list. Now consider the sentence 'There are 4 chairs in the room'. This sentence is true provided that the only marks that combine with 'F_1' and 'R_1x_5' are 'x_1', 'x_2', 'x_3', and 'x_4' (of course, any four marks combining with 'F_1' and 'R_1x_5' will do). In that sense, the numeral does not stand for a mark. These sentences are true or false contingently; their truth-value depends on "how it is in the world." This dependence on the world is reflected in the fact that their truth-values depend on which strings are on the list. A sentence then is contingently true or false if the strings on the list determine its truth-value. It is necessarily true or false if its truth-value is independent of which strings are on the list.

At this point, I will explicate Wittgenstein's atomism, the position that in the world there are only atomic facts. Not all ordinary English sentences take the form of subject-predicate or relation sentences. Some are compounds containing connectives such as 'and', 'or', and 'not'. Wittgenstein argues that the list, the world, determines the truth-values of these ordinary language sentences. The molecular sentence 'Harry is not a chair' is true if the sentence 'Harry is a chair' is false. That is, it is true when 'Harry is a chair' is not correlated to any string on the list. 'Harry' was correlated with 'x_2' and 'is a chair' with 'F_1'. Thus, 'Harry is a chair' is true provided that 'F_1x_2' is a string on the list, and 'Harry is not a chair' is true otherwise. For another example, consider 'Harry and John are chairs'. This sentence is true

---

[2]For purposes of the example, assume that every English proper name has a unique referent.

if the sentences 'Harry is a chair' and 'John is a chair' are both corre-
lated to strings on the list. The point is that the logical connectives
do not correlate to marks on the list.[3] Thus, the atomic sentences
completely determine the truth-values of molecular sentences.

Wittgenstein argues that the quantifiers are to be analyzed as
molecular sentences. Specifically, 'all' is construed as a conjunction
of sentences and 'some' as a disjunction of those sentences. Conse-
quently, Wittgenstein believes that the list determines the truth-values
of quantified sentences. Consider a person deciding whether the sen-
tence 'All chairs in the room are blue' is true in the world presented
by the above example. Say there is one more mark on the list 'F_2'
standing for being blue. The sentence is true if 'x_1', 'x_2', 'x_3', and 'x_4'
are (i) on the list, (ii) combined in strings with 'F_1', 'F_2', and 'Rx_5',
and (iii) are the *only* marks so combined.

Some object to this treatment of the quantifiers on the grounds that
in specifying which sentences belong to the collection to be conjoined
or disjoined, Wittgenstein is implicitly appealing to facts other than
those on the list. Those who object imagine a person searching the
list for a string standing for the *absence* of chairs in the room or for
a further 'x_i' combined with 'F_1' and 'Rx_5' on the list. Since no
such string is on the list, Wittgenstein's opponents conclude that the
account is defective. They demand the addition of a closure clause on
the list saying that the list is complete or that these are all the facts.
According to them, the closure clause is necessary to determine the
truth-values of the quantified sentences.

This argument rests on a confusion. Above, I said of the list that it
is complete, but did not provide for this completeness on the list itself.
This is as it ought to be. It may be the case that one cannot specify
the world or ascertain the truth of a quantified sentence without using
a closure clause in the language. In everyday talk, closure clauses are
necessary, but even there they do not speak to anything in the world,
but rather to one's diligence in examining it. The closure clause itself
ought not to be on the list, but said of the list. Wittgenstein says
that "the world is determined by the facts, and by these being all the
facts." Yet, the world is merely "the totality of all facts"[4] and does

---

[3] *Tractatus*, 4.0312.
[4] Ibid., 1.1-1.11.

not include the additional general fact that these are all the facts.

To illustrate the need for closure clauses in the language even if they do not represent anything in the world, imagine a person looking on the list to see if an enumeration of chairs exhausts the collection of chairs in the room. This person knows that the list is complete, though the completeness is not itself represented on the list. Her enumeration contains the words 'John', 'Harry', and 'Eric', which are correlated to marks 'x_1', 'x_2', and 'x_3'. If this enumeration does not exhaust those on the list that are combined into strings with 'F_1' and 'Rx_5', then there will be another mark on the list so combined. Otherwise, there will be no such marks. However, there does not need to be a further fact that there is another such mark nor that there are no more such marks. Of course, the observer is idealized in that she sees the world *sub specie aeternitatis*. When the world is open to immediate inspection, there is no question of whether some enumeration is exhaustive or not. In ordinary life, however, one does not have the same power to survey the world and so needs to investigate whether an enumeration is exhaustive. I said above that the closure clause is a comment about the list. This was misleading. The closure clause is a comment about the observer, not the world. Consequently, I take the list as being sufficient to determine the truth-value of quantified sentences. Just as when one specifies the domain for a model, one does not include in the domain some entity marking that it is exhaustive, for specifying the domain is specifying its exhaustion. As Wittgenstein expresses it, "When the objects are given, therewith all the objects are given".[5] Applying this principle to the particular case, that certain proper names exhaust the predicate 'F_1' and relation 'Rx_5' shows itself on the list and is not a further fact.

I turn now to the truth makers for α sentences. The remarks on mathematics in the *Tractatus* point to the idea that the numerals in the α's do not stand for entities or, in my schema, marks on the list; in fact, one checks whether α sentences are true or false in a manner roughly similar to the way one checks whether a molecular sentence is true or false. Ontologically speaking, Wittgenstein holds that there are no numbers, just as he holds that there are no logical objects. Rather, how many things there are of a certain kind in a "container"

---

[5]Ibid., 5.524.

shows itself. I will explain this view and give some textual evidence to support it.

To elaborate on what makes an α true, I will discuss the standard for applying a numeral to a collection. By 'collection' I mean the objects satisfying both 'in C' and 'F'. I will treat the statement that n applies to the collection as equivalent to the sentence 'There are n F's in C'. I am not taking a metaphysical position on collections or attributing one to Wittgenstein but using the term 'collection' only to talk about the truth maker for an α.[6] To make this talk smoother, one could cast the α sentences as 'The F's in C are n in number'.

I am introducing standards to talk through the truth makers for ordinary language sentences. A standard is associated with a term. It is an object against which one compares other objects to see if a term applies. When the standard and that to which the standard is compared match, one applies a term to the object under consideration. For example, let the standard for the term 'red' be a colored swatch. 'Red' is applied to something if it is the same color as the swatch. In this case, the object must have something in common with the standard in the ontologically strong sense to match it. The thing they must share is a property. So there is literally something in the standard for 'red' that is also in the thing to which we apply the term 'red'. I speak loosely and do not want to imply that the term applies to its standard. A person commenting on the object that another uses as a standard has a different set of standards.[7]

For another example, the standard for the term 'chair' could be a picture of a chair, and the matching relation is that the picture and the thing to which one applies it have the same look. In this case, of course, I imagine that there is only one kind of chair. Now consider the standard for '3' or any other numeral. The standard could be a piece of paper with marks. The standard matches the collection when the marks on the standard are one-to-one with the objects in the collection. This implies that numbers show themselves; there is no

---

[6]See *Tractatus*, 6.031.

[7]I have been considering an observer looking at the world from outside of it, or *sub specie aeternitatis*. If I am justified in considering this observer at all, I am also justified in imagining her as having standards which are themselves outside of the world. The observer uses these standards to apply terms to objects in the world. Yet, she does not apply the terms to her standards.

mark on the list corresponding to the numeral.

I will now recast my talk of the world making a sentence true in terms of objects matching the standards for the terms in the sentence. I will explain how to use standards to talk of a contingent subject-predicate or relational sentence being made-true. Consider a sentence of the former kind, for example 'a is red'. On the traditional account, the sentence is true when one finds the particular indicated by 'a' and checks to see if it satisfies the predicate 'is red'. It satisfies the predicate provided that it matches the standard for 'red'. Similarly, one checks the truth-value of a relation sentence, for example 'John is in the room', by seeing if the pair of terms matches the standard for the relation-term 'is in'. This analysis can be extended to other contingent sentences, such as the α's, which are not correlated to any single strings on the list. An α sentence, recast as 'The number of F's in G is $\underline{n}$', is true when the "collection" of F's in G matches the standard for '$\underline{n}$'.

I now will shift back to discuss this in terms of the list. Imagine that someone wants to decide which numeral to apply to the chairs in the room. She points to 'x_1' and says '1'. Then to 'x_2'and says '2'. She points to each further 'x_i' combined with 'F_1' and 'Rx_5' and utters the next numeral until she has pointed to every such 'x_i'. The numeral she utters while pointing at the final 'x_i' is the one she applies to the entire collection. It is clear then that the collection of numerals up to and including '$\underline{n}$' count as the standard for the term '$\underline{n}$' and the matching relations is that the two collections are one-to-one. Thus, the standard for applying each numeral corresponds to its position in the sequence of natural numbers. Paul Benacerraf makes a similar point in "What Numbers Could Not Be." In counting, we take the "elements [of a collection] one by one as we say the numbers one by one." In this way, one establishes that the collection matches the sample for a numeral.[8]

To say the same thing differently, since Wittgenstein holds that the numerals do not stand for objects, the matching relation between a standard for a numeral and a collection cannot be that the collection and the standard share something in the ontologically strong sense. For now, imagine the standard associated with a numeral '$\underline{n}$' to be a

---

[8]Benacerraf, p. 50.

piece of paper with marks on it. Wittgenstein thinks that a collection matches a number standard if the standard and the collection are one-to-one. Two collections, say the F's in C and the G's in D are one-to-one when they can be associated in such a way that for every object that is both F and in C there corresponds exactly one object that is both G and D. So one applies a numeral '$\underline{n}$' to a collection when the collection is one-to-one with the standard for '$\underline{n}$'. In actual practice, the standard for an '$\underline{n}$' is the collection of numerals up to and including '$\underline{n}$' and the matching relation is being one-to-one. The strings on the list determine which objects are in a collection and thereby determine the truth-value of an α. This accounts for the contingency of the α's.

To illustrate what makes two collections one-to-one, consider the world with four chairs posited above, with one further stipulation. Instead of only four proper names in the language, there are eight. The other four, 'Kristen', 'Mary', 'Cynthia', and 'Jenna', stand for 'x_6', 'x_7', 'x_8', and 'x_9' on the list. These are all people in the room, so they each combine with both 'F_2', which is represented by the English predicate 'person', and 'Rx_5' into strings on the list. Finally, there is a spatial relation-term, 'R_1', standing for the relation of x being on y. Imagine that one wants to check to see if the people in the room are one-to-one with the chairs in the room. An obvious way to test for this would be to have each person actually sit in a chair and see if there are any chairs or people left over. If there are chairs left over, this shows itself by there being a chair on which no one is sitting. Similarly, if the chairs and people are one-to-one, then this shows itself as well by every chair on our list having exactly one person on it.

At 4.1272, Wittgenstein hints that a standard for a numeral and a collection match when they are one-to-one where he construes 'there are two objects which...' as '$\exists(x, y)...$'. In this remark, the numeral is eliminated and partially replaced by two variables. The number of variables under the quantifier is the correct number to apply to the collection. The variables quantified over must then be one-to-one with the objects. More to the point, in the *Notebooks* one finds that "what the pseudo-proposition 'There are n things' tries to express" shows itself "in language by the presence of n proper names".[9] So more particularly, one applies '3' to a collection provided that the marks on

---

[9] *Notebooks*, 28.10.14.

the standard for '3' are one-to-one with the objects in that collection.

In *Foundations of Arithmetic*, Frege, like Wittgenstein, holds that a numeral applies to a collection provided that the standard and the collection are one-to-one. However, Frege believes that in virtue of two collections being one-to-one, the collections have something further in common. That a collection is one-to-one with the standard for 'n̲' is a justification for asserting 'The F's in C are n̲ in number'; yet, this is not all there is to it. Frege believes that there must be additional marks on the list when two collections are in one-to-one correspondence.

To illustrate what Frege is asserting in terms of the example of the people in the room being one-to-one with the chairs, Frege believes that in addition to 'x_1', 'x_2', 'x_3', and 'x_4' each being combined in strings on the list with 'F_1' and 'x_6', 'x_7', 'x_8', and 'x_9' each being combined in strings with 'F_2' and these being the only proper names combined with 'F_1' and 'F_2', there must be further strings on the list making true the sentence 'The chairs in the room and people in the room are one-to-one'.

Frege takes these further strings to be required to represent the truth makers for the α's. One says that the F's in C are one-to-one with the G's in D when the same 'n̲' is true of both the F's in C and the G's in D. According to him, an α is true only if (i) the standard for the numeral in the α is one-to-one with the collection and (ii) the α is correlated to a string on the list in which the numeral, 'n̲', and the collection are correlated to marks in the string. Frege then is committed to the view that numerals stand for entities whose criterion of identity is that all collections they combine with are one-to-one. He says numbers are "self subsistent objects." As is well known, Frege holds that each 'n̲' stands for a particular set, but this is less significant than the fact that an 'n̲' stands for an entity at all. Further, Frege needs a mark on the list for "the something" to which the number applies. He takes it that the number is attributed to the property, 'F_i'.[10] However, an α should assert something about the objects bearing 'F_i' and not about 'F_i' itself. So, for 'The F's in C are n̲ in number' to be true, there must be a string on the list, '$x_i R_j x_k$' such that 'the F's in C' stands for 'x_i' and 'n' stands for 'x_k'. For Frege, then, the α's are in effect atomic.

---

[10] *Foundations*, pp. 68-69.

Of course, he denies that the '<u>n</u>' in an α represents anything which can be sensed. One cannot, he says, have any sort of image or sense presentation of, for example, three-ness. Someone imagining or picturing three horses, he would say, does not picture three-ness, but only horses. No element corresponding to '3' in the α appears in the mental representation; but Frege insists that such an element exists in the world. He concludes that numbers are not the right kind of entity to be shown in a picture. Wittgenstein would say that a picture or mental image represents that there are three horses because the things representing horses in the picture or mental image are one-to-one with the standard for the numeral '3'. For Wittgenstein, no entity corresponds to '3' in either the mental image or the world.

This case shows the parsimony of Wittgenstein's position. Frege recognizes that being one-to-one with a standard is a sufficient and necessary condition for an α to be true but does not think that this is all there is to the matter. As a result, he includes entities in his ontology that seem redundant. If being one-to-one with a standard is a sufficient and necessary condition for a numeral to apply to a collection, then there is no need of further sentences on the list to represent that collections are one-to-one.

Frege goes wrong because he thinks that to match a standard to an object requires something to be identical in the standard and what it matches. In the simplest example, one compares a color-swatch to an object, saying that they match when the color in the object is the same thing as the color of the swatch. Frege thinks of being one-to-one as justifying a number assertion, but not exhausting its meaning. He tries to illustrate this by considering someone deciding the direction of line Γ, by comparing it to standard lines. Say she is comparing Γ with line Δ, which is the standard for direction D. The direction of Γ is D provided that Γ matches Δ. Two lines match when they are in the same direction. Frege, then, believes that "matching" a standard line is a justification for saying that D is the direction of a line. In this case, that Γ is in direction D is represented on the list.[11]

Yet, this example is not analogous to the case of comparing the standard for a numeral to a collection. In the example, the matching relation is same direction. That is, Γ matches Δ if both Γ and Δ

---

[11] *Foundations*, pp. 73-77.

have the property of being in a certain direction and those directions are the same. Thus, the matching relation between standard and collection involves sharing a property. But being one-to-one does not involve sharing "properties." That two collections are one-to-one is not represented by a string on the list. Once this is appreciated, the motive for assuming that an α designates further strings on the list vanishes.

BRYAN PICKEL
University of Texas, Austin

Bibliography

[1] P. Benacerraf. What Numbers Could Not Be. *Philosophical Review*, 74:47–73, 1965.

[2] G. Frege. *The Foundations of Arithmetic*. trans. by J. L. Austin. Northwestern University Press, Evanston, 1980.

[3] L. Wittgenstein. *Notebooks, 1914-1916*. G. E. M. Anscombe and G. H. Von Wright, editors, and trans. by G. H. Von Wright. University of Chicago Press, Chicago, 1984.

[4] L. Wittgenstein. *Tractatus Logico-Philosophicus*. trans. by C. K. Ogden. Routledge, New York, 2002.

A Defense of Eliminativism

Eliminativism is the view that ontological simples[1] are the only extant entities. Simples, on this view, are defined as material or physical objects without component parts. According to the eliminativist, objects that can be broken down into component parts are nothing more than arrangements of simples, and do not, strictly speaking, exist. This is a very strong claim considering that nearly everything in the world around us, including ourselves, is an object that is composed of parts.

Although eliminativism solves various philosophical puzzles, it engenders many problems of its own. Not surprisingly, most of these problems have as much to do with the philosophy of language as they have to do with metaphysics. This is because the debate over the plausibility of eliminativism is a symptom of a much larger debate

---

[1]Of course, what exactly these simples are that do exist is a subject of debate, the resolution of which may ultimately lie in the hands of science. After all, science has made great progress over the centuries in discovering smaller and smaller particles that plausibly make everything else up. It may well turn out, however, that the hypothesis that there are ontological simples at all is false. It may be that everything can be reduced to something else. The eliminativist need not wait for science to finally make this assessment, however, since his theory does not rely on knowing what these simples are. Analogies can be, and are, readily made. Instead of hypothesizing that quarks or gluons ultimately make up a house, the eliminativist can make his arguments just as well by speaking of the bricks that make it up as the ontological simples. The essential point is to establish that the house does not exist in the eliminativist ontology, not to establish exactly what does. As van Inwagen puts it, whatever the simples are, they act as the furniture of the world; they are arranged and rearranged in such a way that they make everything else up. Only this 'making up' something does not mean that this something exists.

involving doctrines in the philosophy of language and in metaphysics. On the one hand, we have the view that our ordinary way of speaking about the world of material objects is a result of a largely accurate conception we have of objective reality.[2] The view is that a correct ontology would contain most of the material objects we ordinarily speak about as existing. On the other hand, we have the view that our everyday linguistic practices do not reflect objective reality at all. All we can say is that they are based on our conception of the world and therefore reflect our reality. The view here is that the correct ontology contains far less or far more than what we ordinarily think of as existing. Needless to say, the former view is quite antithetical to eliminativism since eliminativism entails that much of what one normally speaks of as existing doesn't exist.

In this paper I will take up a position on the side of those eliminativists who would argue that our conception of the world does not provide an accurate picture of the way things actually are. My aim is to defend eliminativism against the view that it cannot be true because it goes against much of what we commonsensically believe. As I will argue, our acceptance of eliminativism neither requires our relinquishing our ordinary way of talking about physical objects nor our concession that it is flawed. Moreover, it does not entail that when we speak about ordinary objects we are necessarily saying something false. Instead it simply requires a sensitivity to the difference between what is part of our conceptual scheme and what exists in objective reality.[3] My strategy for making this argument, and thus defending

---

[2]For simplicity sake, throughout the course of this paper I will use the terms 'objective reality' and 'the world' to refer to mind independent reality, even though any objective view of the world would contain us. My intention is to draw a distinction between our conception of the world and the way it might actually be. I am simply assuming that there is such a mind-independent reality without attempting to argue for it.

[3]Making this distinction, it should be noted, does not involve the question of whether truth is relative or objective. The Relativism/Realism debate over the nature of truth, although it may play a part in the debate outlined above (if it is formulated somewhat differently), is irrelevant once the question has focused in on what objects should be included in a *correct* ontology. This is because philosophers who are at all interested in whether eliminativism can plausibly be correct must already think that there is a certain way things are in reality. If so, it seems they must be committed to the view that something is true only if it is in some kind of accordance with that reality.

eliminativism, is to closely examine a criticism of eliminativism as found in some recent work of Eric Olson,[4] and then to show how Peter van Inwagen's[5] eliminativistic position regarding material objects stands up to this criticism.

Olson presents a critique of eliminativism[6] by asking the following question of the eliminativist: *How can the eliminativist maintain that sentences referring to collections of simples as existing objects are true in non-philosophical contexts without having to agree that this truth is not at all based on an agreement between the sentence and the way things really are?* In other words, Olson suggests that the view that ordinary sentences can be strictly true (in the Realist sense) is incompatible with eliminativism. This, on his view, leaves the eliminativist in the unsatisfactory position of having to admit that ordinary sentences about material objects are never strictly true.[7] In what follows I will argue that van Inwagen presents an eliminativistic account regarding non-living material objects that resists Olson's conclusions. His account does this by outlining how it is that we actually do say strictly true things when speaking about material objects.

Van Inwagen acknowledges that many of the problems eliminativism gives rise to involve the relation between our conceptual scheme and objective reality, when he says:

---

[4]See Olson 1997: *The Human Animal: Personal Identity Without Psychology* (HA), Ch. 7.

[5]As found in *Material Beings* (MB). Most of the material relevant to this paper can be found in sections 10, 11 and 13.

[6]On page 72-73 of *Material Beings*, van Inwagen refers to the kind of eliminativism he espouses as Nihilism. Quite interestingly, van Inwagen's Nihilism holds that organisms that constitute a life also exist along with simples. For the purposes of this paper this latter amendment will not be of much concern, as it makes no difference to the argument at hand. Also of interest is the fact that Olson's paper specifically challenges eliminativists who, unlike van Inwagen, claim that people do not exist. I think it is partly because Olson does not think of his criticism as being directed at van Inwagen, that he does not notice the solution van Inwagen's position offers against his criticism. Of course, the claim might be made that Olson's criticism is only directed at theories that deny the existence of persons. I see no reason to think this is the case, especially considering his comments on eliminativism in his article *Composition and Coincidence* (cf. note 6). But even if it is, I see no reason why the following account could not be modified to work against this more limited form of criticism.

[7]The other option, *prima facie*, is to accept some form of relativism. But for reasons already given this option is ruled out.

> For any philosopher who denies what practically everyone
> else believes is, so far as I can see, adopting a position
> according to which the human capacity for knowing the
> truth about things is radically defective....It is far from
> obvious, however, that it is a matter of universal belief that
> there are chairs....*It is difficult to settle such questions in
> part because there are a lot of things one might express by
> uttering philosophical sentences like 'there are chairs'.*[8]

Olson's criticism of eliminativism is based on just the linguistic is-
sues van Inwagen draws attention to above. Olson wants to maintain
the epistemological position that our "capacity for knowing the truth
about things" is not radically defective. However, on Olson's view
there is no getting around the fact that eliminativism entails that ev-
ery sentence we utter involving objects of the standard ontology is false
or non-sensical, simply because if eliminativism is true these objects
do not exist.[9] The only course for the eliminativist, if he is to avoid
having to accept a form of epistemological skepticism by saying that
we can't know the truth about reality because our capacity for doing
so is defective, is to establish that eliminativism is, to some degree,
linguistically compatible with our ordinary conceptual scheme.[10] In
other words, the eliminativist is going to need to show that our ordi-
nary way of talking about the world can be maintained even though
very few of the objects to which we refer exist at all.[11]

Short of being able to establish the strict truth of ordinary sen-
tences, all the eliminativist can do, according to Olson, is establish
what *is correct* about sentences referring to material objects. This
involves making the crucial distinction between sentences such as (1)
'there is a chair over there,' from sentences such as (2) 'there is a uni-
corn over there.'[12] After all, if eliminativism is true, both of these
sentences are false, as they both refer to objects that don't exist as

---

[8]MB, p. 103. Emphasis my own.

[9]Olson never makes this his official theory of truth, but I think it is clear enough
that he holds this sort of 'referentialist' position.

[10]See Olson 1996: 'Composition and Coincidence' (CC), p. 398. Olson writes:
"The eliminativist may adopt a semantic theory...that brings her view into line
with things that every sane human believes."

[11]CC, p. 398.

[12]HA, p. 156.

being in the vicinity of the speaker. The relevant facts that need to obtain to make these sentences true do not obtain. Moreover, these facts fail to obtain for the same reason: the objects in question do not exist.

Olson points out that the eliminativist can show why it is not false to say (1) in the same way as it would be to say (2) by paraphrasing sentences such as (1) to read: (1') 'There is a collection of simples arranged chair-wise over there.'[13] (1') has the specific advantage, for the eliminativist, of not referring to any thing that he does not accept as an existing object. Because of this (1') if true, (i.e., if there is actually a collection of simples arranged chair-wise present) can be considered strictly true in a way that (1) cannot. Furthermore, because it can be paraphrased by a sentence that is strictly true, (1) is importantly different from (2).

Because sentences such as (1) can be paraphrased by sentences that are strictly true and sentences such as (2) cannot, Olson characterizes the former as 'harmlessly false,' or more euphemistically, 'appropriately true,' whereas sentences like (2) are always flat out false and presumably, harmfully so. The *appropriateness* of these otherwise false sentences establishes the 'compatibility' between eliminativism and our ordinary conceptual scheme. In other words, even though sentences referring to objects of the standard ontology are strictly false, because such sentences can be paraphrased by sentences that are strictly true, the eliminativist can maintain that sentences such as (1) say *something* correct about reality.

Olson's point about the appropriateness of (1) over that of (2) is made clear if we consider what it is to believe in the referents of these two types of sentences as existing. A belief in chairs is something that seems, at least in ordinary life, rather advantageous. This belief, if eliminativism is true, is a harmlessly mistaken belief,[14] having an element of 'cognitive virtue' that a belief in unicorns or ghosts lacks. Better put, a belief in chairs is only harmful in certain philosophical

---

[13]See note 9. Olson continues: "The way to do this is to show how to paraphrase...". Also see HA p. 156, "What makes believing in people different from believing in unicorns? The usual way of answering challenges like this is to try to paraphrase those sentences...into sentences that do not entail that there are people."

[14]HA, p. 155.

contexts, whereas a belief in elves, dragons, or that taxes don't really need to be paid, is quite often harmful in ordinary non-philosophical contexts. Olson suggests that the ordinary belief in chairs is analogous to a mathematician's belief in prime numbers.[15] Even if prime numbers do not, strictly speaking, exist, this doesn't mean that the mathematician's belief in prime numbers is a harmful mistake, at least not while doing mathematics. If, however, the mathematician believed in even prime numbers other than 2, such a belief would be like the ordinary belief that taxes don't ever need to be paid. It would surely make the mathematician lose his job for the same reason it would make the tax evader lose his shirt. Olson's point is that such beliefs lack the cognitive virtue that other equally mistaken beliefs have. The upshot of Olson's position is that the most the eliminativist can do is to say that such sentences can be used and considered appropriately true. But this 'truth' is not in accordance with objective reality since the objects these sentences refer to don't really exist. Therefore, the eliminativist, although able to maintain that our ability to speak about the world is not radically defective because he can employ paraphrases, is nevertheless forced to concede that our ordinary ways of talking do not express objective truth.

Van Inwagen also holds the view that sentences such as (1) are importantly different from those such as (2) and that it is critical for the eliminativist to establish this:

> If you were to tell the ordinary man that I thought there were no chairs he would probably think I was mad....He would think I regarded utterers of [sentences such as 'there are two valuable chairs in the next room'] as he (perhaps) regards utterers of the sentence 'There are two horrible ghosts in the next house.' But my assertion, (and his, and yours) that there are no ghosts is not like my assertion [while doing metaphysics, that there are no chairs. My assertion in this context that there are no chairs] is not meant to deny that reports of [chairs] are reports of a real and unified set of phenomena. My assertion that there are no ghosts is meant to deny that reports of ghosts are

---

[15]HA, p. 156.

reports of a real and unified set of phenomena.[16]

On van Inwagen's view the assertion that there are chairs is not like the assertion that there are ghosts because the assertion that there are chairs is a report of a *real and unified set of phenomena*. For van Inwagen these phenomena can be accepted in an eliminativist ontology. The question is, of course, whether the unified set of phenomena that makes (1) different from (2) means that (1) is more than merely 'appropriately true' in Olson's sense.

Both van Inwagen and Olson, on the above interpretation, characterize the main problem facing the eliminativist as a problem involving the compatibility of the propositions expressed by our ordinary sentences with the correct ontology. As a further example of this van Inwagen and Olson both agree that if eliminativism is true, even though certain objects do not strictly speaking exist, the eliminativist might want to say that certain identity sentences such as (3) 'that chair over there is the same one that was there a year ago' are appropriately true in non-philosophical contexts. After all, one of the reasons a philosopher like van Inwagen is attracted to eliminativism in the first place is that it allows him to say that a certain virtual chair[17] at a certain time (t) is the same 'chair' at another time (t*) without becoming engulfed in *Ship of Theseus* type problems concerning the identity of objects, especially artifacts, over time.[18]

That this feature is attractive to van Inwagen is made clear when he says that "if we follow this rule [i.e., if we translate ordinary sentences into paraphrases containing no reference to artifacts] we shall be able to formulate no philosophical questions about the identities of artifacts at all."[19] This avoidance of all philosophical questions about the identities of artifacts, I take it, boils down the avoidance of such questions as: "In virtue of what is something strictly identical with itself when there is a change of component parts involved?" These

---

[16]MB, p. 107.

[17]For the sake of clarity, van Inwagen calls those objects which are not included in the eliminativist's ontology 'virtual objects' and suggests that the eliminativist may want to assert an identity sentence (MB, p. 125). Olson points out the same thing in HA on p. 157, and CC on p. 399.

[18]MB, p. 128. Van Inwagen also alludes to this advantage in MB, p. 140, when he writes that it would be pleasant to avoid such tendencies.

[19]MB, p. 130. The square brackets indicate modifications I have made to van Inwagen's texts to make it appropriate to the issues here.

kinds of problems are avoided because if neither the chair at t nor the chair at t^* exists then the former can be strictly identified with the latter only if the simples in a chair-wise arrangement at t are exactly the same simples as those in a chair-wise arrangement at t^*. But this is what engenders *Ship of Theseus* type problems; there is never such strict identity of particles because the parts of a 'composite object' invariably undergo change. Thus, if there is to be strict identity it must be a relationship between an object at t and an object at t^* as they exist over and above their parts. Well, according to eliminativism no such objects exist. Therefore, there can be no such relationship between objects. With eliminativism the problems of strict identity over time are solved "in the same way that the atheist solves the problem of the holy trinity."[20] If the eliminativist is to ask anything in this regard he will only be able to ask in virtue of what can a 'virtual artifact' be said to persist over time.

However, in order to say that an identity sentence like (3) is importantly different from a similar identity sentence about unicorns or ghosts, the eliminativist must be able to first say that the identity sentence, as van Inwagen puts it, "reports a real and unified set of phenomena."[21] But if the eliminativist is to do this he must be able to say what it is in virtue of which the *virtual object* persists. Van Inwagen puts the problem like this:

> But we do, in the ordinary business of life, utter identity sentences that refer to artifacts. And in uttering these sentences we sometimes say things that are true and sometimes say things that are false. If we say, (4) 'This is the house that Jack built,' and it is the house that Jack built, then we're right, and if it is the house that Jill built, then we're wrong. How would you account for the difference?[22]

That is, we must be able to say what it is about Jack's house that makes it his house rather than Jill's or someone else's even if it is only a *virtual* house, otherwise we can't say that the identity sentence

---

[20] As Olson points out in HA, p. 154.

[21] MB, p. 107. See quote above. As I take it this phenomenon is not to be understood in the phenomenalist sense, but is real in the sense that it exists in mind-independent reality.

[22] MB, p.129.

describes any real phenomena at all. Olson, for his part, asks what this real phenomenon involved in appropriately true identity sentences might possibly consist in:

> The eliminativist may want to say that some statements about the identity [of artifacts] are true, or at least not misleading. Thus we can try to find paraphrases for those sentences that are strictly true...but what relations among particles...at different times must be true (or appropriate) to say that a single [artifact] exists at two different times?[23]

As might be suspected, Olson characterizes whatever real phenomenon there might be for the eliminativist as a relation between collections of simples, rather than a relation between objects. Again, whatever this relation is, it could not be one of strict identity unless all the particles at t and t^* were numerically the same. Moreover, van Inwagen's account gives the initial impression that he has a similar identity relation in mind as what constitutes the persistence conditions of a virtual object. This seems to be the case because of what van Inwagen presents as an answer to the question of what the persistence conditions of a virtual artifact are, if that artifact does in fact persist over time. The house is Jack's, rather than Jill's, if the present simples that make up the house belong to the same *history of maintenance* as the simples that have composed Jack's house all along. The spatio-temporal path along which the Ship of Theseus was regularly maintained—so that it could continue to function as a ship—is the path we tend to think the ship persists along, rather than along some other path (e.g., along a path traced by one of the planks until the point when all of the original planks were put back together again). In other words, it is because an artifact has a history of maintenance, according to van Inwagen, that we tend to see it as persisting over time as a single artifact despite a change of its component parts.[24] A formulation of van Inwagen's History of Maintenance Theory, as I read it, would be this:

> A virtual object persists along a certain spatio-temporal path if at every point along that path the simples that

---

[23]HA, p. 157.
[24]MB, p. 135.

make up the virtual object are part of the same history of
maintenance as the simples that make up the virtual object
at any other time, where the history of maintenance is
simply a history of replacement of simples or modification
of arrangements of simples.

On van Inwagen's view this "sharing a history of maintenance" is a
fact that needs be expressed in a paraphrase of (4), so that only refer-
ence to objects that exist is made, in order to establish that having a
history of maintenance is a real phenomenon. Again, establishing this
is necessary in order to say what is correct about (4) by maintaining
a distinction between (4) and an identity sentence about, say, ghosts.
Van Inwagen's paraphrase, accounting for the identity of the virtual
house, would look like this:

(5) There are objects [simples] arranged house-wise here
now, and these objects are the current objects of the his-
tory of this house-wise arrangement that has continued
without interruption for three hundred years.

However, it is not immediately clear what these persistence condi-
tions, or this "history" might consist in. Therefore, there will invari-
ably be several different paraphrases the eliminativist could come up
with, each displaying a real phenomenon that might be characterized
as the relationship the persistence of the virtual object consists in. For
example, we could have both:

(5') There is a collection of simples arranged house-wise
here now and the simples in that arrangement are a part
of a causal history that has extended back to a point 300
years ago when a similar collection of simples were similarly
arranged.[25]

and

(5") There is a collection of simples arranged house-wise
here now and for the past 300 years there has been, in this
very spatial location, a spatially continuous collection of
simples very similarly arranged.

---

[25]See CC p. 399 for Olson's formulation of a paraphrase containing a similar
'causal theory.'

Both of (5') and (5") are possible paraphrases of (4), and both embody a theory of a certain house's persistence conditions. We can call these theories the *Causal Theory* and the *Block Theory* respectively. Moreover, each might be considered an alternative to van Inwagen's *History of Maintenance Theory* in order to answer the question in virtue of what we can say that a virtual object persists. Therefore, each of these paraphrases might actually stand in competition with (5) as the correct paraphrase of (4). Of course much elaboration on what these theories actually say and entail is necessary before they can be considered viable alternatives to the *History of Maintenance Theory*, and thus such considerations are necessary before we can say that (5') and (5") are viable alternatives to (5). However, even if such considerations lead to the conclusion that these theories are not plausible alternatives to (5), surely other theories can be offered that are. The point is that because there are competing theories available as to what the relationship is between the virtual house at t and the one at t^*, there are competing paraphrases of (4).

If, for the present purposes, we assume that these theories are viable alternatives to van Inwagen's *History of Maintenance Theory*, then the question is how an eliminativist would choose one of these paraphrases over the others. The only recourse for the eliminativist is to decide which theory expresses the correct metaphysical relationship between virtual objects as all three paraphrases merely describe the same fact that there are simples arranged house-wise at t and there are simples arranged house-wise at t^*. In other words, on both Olson's and van Inwagen's views, the real phenomenon accounting for the persistence of the house must be some kind of metaphysical relationship and the eliminativist must determine which paraphrase best describes this relationship.

On Olson's view the only way the eliminativist can determine which paraphrase best describes the relationship that accounts for the persistence of a virtual object is to ask which paraphrase comes closest in meaning to the meaning of the original sentence (4). This is a linguistic question, rather than a metaphysical one. Olson sees this as a perfect example of the shortcomings of eliminativism. The eliminativist thinks he is asking a question about the way the world really is when formulating his theory about the persistence of a virtual object, but he is really only asking what it *means to say* that the 'object'

persists.

This conclusion is particularly problematic for van Inwagen because on his account a paraphrase describing the persistence conditions of a virtual object is either true or false depending on some fact in the world, rather than merely on what it means to say that the virtual object persists. In other words, on van Inwagen's view it is still a straightforwardly metaphysical question regarding what it is in virtue of which a collection of simples persists over time. On Olson's view, however, instead of asking a metaphysical question about a fact in the world, the eliminativist can only ask which paraphrase best describes the relationship between two virtual objects over time.

As stated at the outset, I think van Inwagen's account plausibly succeeds in showing that our ordinary way of speaking about the world is not radically defective, even if eliminativism is true. Our ordinary way of speaking is not defective because questions regarding virtual objects can remain questions about objective reality. For example, we can ask what the persistence conditions of a virtual object are and offer a theory as an answer to this question. However, this position is only tenable, given Olson's criticism, because of a certain claim van Inwagen makes in the philosophy of language. The claim is that *ordinary sentences are ontologically innocent.*[26]

The thought behind this striking claim is that what we say and believe can accord with certain ontological facts (that there are only simples arranged in certain ways) even though those facts contradict the facts asserted by the sentence (that there are houses, etc.) because the truth of ordinary sentences does not entail that the objects they refer to exist. In other words, even if there are, strictly speaking, no facts about houses (and thus no facts according to which houses persist), there still are facts in virtue of which the proposition expressed by a sentence referring to houses can be strictly true. What makes the proposition expressed by a sentence true, on this view, is the same set of facts in virtue of which the proposition expressed by the correct paraphrase is true. These facts are explicitly given by the paraphrase. That is, the fact that there is a collection of particles arranged chair-wise present is what makes the original sentence (1) and

---

[26]MB, p. 113. Van Inwagen asserts that the original sentence "is neutral with respect to the question of whether there are objects that fit exactly into chair receptacles."

the paraphrase (1') true.

   If this is right then Olson's criticism is off the mark because it means there is a set of facts that makes (4) strictly true, not just harmlessly false or appropriate. When faced with a choice between (5), (5') and (5") the eliminativist can make his decision by considering which theory embodied in each of these paraphrases best expresses these facts. Because identity sentences are true in virtue of certain facts about the world, rather than being merely appropriate based on their meaning, which paraphrase best expresses these facts is a metaphysical question concerning what the objective facts are.

   Of course, whether or not we can accept van Inwagen's account and thus have reason to reject Olson's criticism depends on the plausibility of the claim that ordinary sentences are ontologically innocent or, more specifically, non-referential in that their truth does not depend on the existence of the objects they refer to. This becomes even more clear if we take the critical line that, as has conventionally been thought, the only fact that could make (3) true is the fact that there is a chair over there. That is, the proposition expressed by the sentence token 'there is a chair there' is true because of the fact that the chair that is being referred to exists. If this is the correct way to understand the truth of sentences, and if eliminativism is true, then, to be sure, (5) cannot be true. This is surely how Olson sees matters. Unfortunately, van Inwagen offers little in the way of a defense for the ontological innocence claim. However, I believe the following could be reasonably offered as such a defense: Even though the proposition that there is a chair over there is true only if there is what *is called* a chair over there, what makes the proposition true is not that this 'chair' exists, but again, that certain facts in the world hold.

   More formally I am proposing the following argument:

   1) The proposition expressed by (a token of) a sentence such as (3) is true iff certain facts hold.
   2) The facts that must hold for the proposition expressed by (3) to be true need not include the fact that (3).

   Therefore,
   3) The facts that make the proposition expressed by (3) true do not entail that there is a chair present.

   Whether or not van Inwagen would endorse something like the

above argument as being behind his 'neutrality of ordinary sentences' claim is, of course, debatable. However, he offers a useful analogy which I think can be taken as evidence that he has something along these lines in mind. He suggests a sentence such as (6) 'the sun moved behind the elms' (where (6) is analogous to a sentence referring to a chair uttered in an everyday context) is a perfectly true sentence, even though the sun does not move at all. However, because the sun doesn't move, (6) might be better *expressed* by (7) 'Because of the rotation of the earth, the sun ceased to be visible as a clear view of it became obscured by a group of elms.'[27] (7) of course, is analogous to a paraphrase such as (5). (7) certainly might be true as well, and if so it's true because the facts it describes obtain. Van Inwagen's claim is that the obtaining of these facts is also what makes (6) true. Because the truth-conditions of (6) and (7) (or the facts that must hold for (6) and (7) to be true) are the same, (6) is compatible with the Copernican Hypothesis. Analogously, because the truth conditions of (4) are the same as that of (5), (4) is compatible with the eliminativist ontology.

What might be made of this account? *Prima facie*, there seems to be one major reason to reject it: A proposition either expresses a truth or it doesn't. The above view, the objection might run, wants to maintain that a proposition can fail to express a truth, because the object it is about doesn't exist, but still be true. This seems absurd, unless we are dealing with two different concepts of truth.

The mistake made with the above objection involves a confusion of grammatical or conceptual facts with the facts about the arrangement of collections of simples, or facts about the world. The latter, as established above, are what must hold for the proposition expressed by a sentence to be true. The former, the grammatical facts, however, are what might make the proposition expressed by the utterance of a sentence logically true in certain contexts. That is, the proposition that there is a chair there is logically true iff there is what *we call* a chair there. But this *is* a matter of logic and of language. The concept of truth, I think, need not be so narrowly defined. In other words the proposition that there is a chair over there might be true not because whatever we are referring to happens to be *called* a chair, but because what we are referring to incidentally as a chair happens to exist as a

---

[27]MB, p. 113.

certain arrangement of simples. True enough this may, in some cases, make it very difficult to say why certain propositions are true, as it requires, for example, coming up with the right theories about what holds certain collections of simples together as collections, as well as theories about how collections of simples persist over time. However, in my opinion, this is a relatively small price to pay, considering that this view provides us with the very useful distinction between logical and metaphysical truth. This distinction is useful because it is the confusion between logical and metaphysical truth that keeps us tied to an ontology that is almost entirely based on the words we use. Doing away with this confusion, along with doing away with epistemological skepticism, affords us the opportunity to refine our ontology based on science rather than resting it on convention and practice, without having to continually question our commonsensical view of the world around us.

In the final analysis, the major obstacle facing an eliminativistic ontology, can be turned into its greatest virtue. That is, the argument that if eliminativism is true, then most of the sentences we utter in non-philosophical contexts are false, forces the eliminativist to accept the progressive distinction between our logical conception of the world and the *truth about the world*. I believe this fundamental dichotomy in the notion of truth underlies van Inwagen's 'ontological innocence' claim.

If van Inwagen's view is correct along with the small clarifications I have made to shore it up against Olson's criticism, I think we can explore the possibility that the eliminativist ontology might be the correct one without being forced to concede that it is only while paraphrasing ordinary sentences, due to the pressures of our metaphysical discourse, that we will be able to say anything strictly true. This is because, as I think van Inwagen's account makes clear, our ordinary way of calling chairs 'chairs,' and houses 'houses,' not only form sentences that express logical truths, but also sentences that are strictly true in virtue of certain ontological facts. More importantly, these facts are *not* the sort that make further metaphysical inquiry into the persistence or modification collections of simples impossible, turning our investigation back upon our logic and language once again. Instead, they are facts that allow for further metaphysical investigation into how a collection of simples might maintain itself, be maintained,

modify itself, or be modified *all the while as a collection*, so that we can look at our world of 'objects' persisting through time in a more appropriate way, or more appropriately, a more truthful way.

JASON TIPPETTS
University of Washington, Seattle

Bibliography

[1] E. T. Olson. Composition and Coincidence. *Pacific Philosophical Quarterly*, 77:374–403, 1996.

[2] E. T. Olson. *The Human Animal: Personal Identity Without Psychology*. Oxford University Press, Oxford, 1997.

[3] P. van Inwagen. *Material Beings*. Cornell University Press, New York, 1990.