

Chapter 12

Replies

Peter van Inwagen

12.1 Rational Peer Disagreement: Reply to F. Bögner, T. Meyer, K. Schnieder and M. Seidel

It would be pleasant if we could resolve all the problems and paradoxes and embarrassments with which “revealed peer disagreement” presents our discipline simply by accepting the following general epistemological principle:

(SE) It is possible for there to be a body of evidence that is sufficient to warrant belief that a certain proposition is true *and* sufficient to warrant belief that that proposition is false.

Consider, for example, the proposition that free will is incompatible with determinism. On the basis of careful consideration a certain body of evidence, I believe that this proposition is true. On the basis of equally careful consideration of that very same body of evidence, David Lewis believed that it was false. But if (SE) is true, it is at least possible that both our beliefs were warranted – that the evidence each of us had was *sufficient* to warrant his belief.

In “Listening to Clifford’s Ghost” (van Inwagen 2009a), I maintained that if (SE) is true, then careful attention to evidence that is sufficiently strong to warrant belief in a proposition does not “track the truth.” I am often tempted to accept (SE), I said, but I never do, because on every such occasion the shade of W. K. Clifford attends me and whispers the following argument in my ear (a paraphrase of van Inwagen 2009a, 27):

If you believe that a certain proposition is true and Lewis believes that it is false, and if these beliefs are based on or grounded in exactly the *same* body of evidence, then your evidence does not direct you towards the truth of that proposition and away from its falsity. (For if it did, it would have directed *him* away from its falsity and he would *not* have believed that it was true.) Of course, nothing

P. van Inwagen (✉)

Department of Philosophy, University of Notre Dame, Notre Dame, IN, USA
e-mail: Peter.VanInwagen.1@nd.edu

happens without a cause. *Something* must have led you to that belief. But of all the forces in the human psyche that direct us toward and away from belief in the truth of a proposition, only rational attention to relevant evidence *tracks the truth*. Both experience and reason confirm this. If you ascribe truth to a proposition on the basis of some inner push, some “will to believe,” if I may coin a phrase, that does not track the truth, then your ascription of truth to that proposition is not being guided by the nature of the reality that proposition is *about*. If you could decide what to believe by rolling a die, if that would actually be effective, then, in the matter of the likelihood of your beliefs being true, you might as well form them that way.

The implications for (SE) of Clifford’s *post mortem* argument are obvious. Suppose that I adopt the policy of accepting those and only those propositions for which I have evidence sufficient to warrant the belief that they are true. Well and good – but now suppose (SE) is true. Then that policy is no more likely to lead me to the truth than *this* policy (supposing I were actually psychologically capable of believing things at will): “Whenever I wish to decide whether to believe that a proposition is true or false, I will roll a die; if the die falls 1, 3, or 5, I will believe that it is true, and if the die falls 2, 4, or 6, I will believe that it is false.” That is to say, if (SE) is true, then “having sufficient evidence” does not track the truth. And if that is so, why bother with the time- and energy-intensive task of collecting evidence at all?

The authors, however, contend that there is an incoherency in my reasoning. They contend, first, that I accept the following argument:

- (P6) If it is possible for a person *S* to have sufficient evidence *E* for a true belief *p* and for a person *R* to have sufficient evidence *E* for the false belief *non-p*, then rational attention to evidence would not track the truth.
- (P7) Rational attention to evidence tracks the truth.
- (C4) It is not possible for a person *S* to have sufficient evidence *E* for a true belief *p* and for a person *R* to have sufficient evidence *E* for the false belief *non-p*. (Sect. 2.4.2)

They then say,

If one accepts [this argument] and especially if one accepts its first premise, one should, as we want to argue, accept the following argument as well: [...]

- (P8) If it is possible for a person *S* to have sufficient evidence for a false belief *p*, then rational attention to evidence would not track the truth.
- (P9) It is possible to have sufficient evidence for false beliefs.
- (C5) Rational attention to evidence does not track the truth. (Sect. 2.4.2)

I would not go so far as to say that I have presented an *argument* for (C4) – or against (SE), which is something like the denial of (C4). I should rather say that I have tried to show that (SE) has some consequences that I, at least, find unacceptable. But let that pass. Let us consider the first of the two arguments as they present it. I would accept (P6) only in this revised sense (the only really important aspect of the revision is the addition of the phrase ‘totality of’):

- (P6’) If the following two propositions are consistent:
 - (i) *S* believes that *p*; *E* is the *totality* of the evidence that *S* has that is relevant to the question of the truth-value of *p*; *S* has sufficient evidence for *p*,

- (ii) *R* believes that *not-p*; *E* is the *totality* of the evidence that *R* has that is relevant to the question of the truth-value of *p*; *R* has sufficient evidence for *not-p*,
then rational attention to evidence does not track the truth.¹

Am I therefore committed to accepting the second argument? I am not, for (consistently with my accepting the revised version of the first argument) I reject (P8). Or, rather, since I do not have any workable general definition of “sufficient evidence”, I will make this rather more nuanced statement: I contend that what we all know about the human condition (assuming that anyone knows anything about anything) and the proposition “Whatever, exactly, ‘sufficient evidence’ may be, human beings *sometimes* have sufficient evidence for their beliefs about immediately present physical objects” jointly entail the falsity of (P8).

Suppose Lois looks at a field and sees something that looks exactly like a sheep; it is obvious that *if* Lois has sufficient evidence for the proposition “There is a sheep in that field”, it does *not follow* from its sufficiency that there is a sheep in the field. (What she is seeing might be a cleverly crafted *faux* sheep; she might be a brain in vat or a plaything of the *genius malignus*...). Nevertheless – so I believe, and I expect you do, too – the evidence we typically have that pertains to the properties of immediately present physical objects tracks the truth. Not *infallibly* of course, not in every consistently describable scenario. But how many times in the history of the world has a normal human being in ordinary circumstances believed that he or she was seeing a sheep a couple of meters away in full daylight? – why, uncounted billions; and in what proportion of those cases has that belief been false? 0.0000024%? 0.00000008%? *Some* number like those must be right. So: if anyone ever has sufficient evidence for any belief, it is *possible* to have sufficient evidence for a false belief – nevertheless, (unless almost all of us are radically wrong about the general features of the world – 43% of apparent sheep are *faux* sheep; we are brains in vats; the *genius malignus* is real and plying his trade) rational attention to sufficient evidence tracks the truth (at least in the matter of the presence and properties of nearby physical objects). It *can* fail to deliver the truth but it tracks the truth.

And this is perfectly consistent with the truth of (P6'). There is all the difference in the world between these two cases:

Case 1: Lois believes that there is a sheep before her; she is having the experience a person normally has when he or she has an unobstructed view of a sheep in plain daylight about five meters away; this (together with any other “sheep-relevant” evidence she happens to have – say, the absence of any reports that there are a lot of *faux* sheep in the vicinity) is sufficient to ground or support her

¹The conclusion of the revised argument then becomes: The following two propositions are *inconsistent*,

- (i) *S* believes that *p*; *E* is the *totality* of the evidence that *S* has that is relevant to the question of the truth-value of *p*; *S* has sufficient evidence for *p*.
(ii) *R* believes that *not-p*; *E* is the *totality* of the evidence that *R* has that is relevant to the question of the truth-value of *p*; *R* has sufficient evidence for *not-p*.

belief that there is a sheep before her; in point of fact, however, there is no sheep before her.

Case 2: Lois and her friend Lana are standing beside each other facing in the same direction and examining the contents of a large country field; Lois believes that there is a clearly visible sheep about five meters from her and Lana; Lana believes that there is no sheep in the field at all; the totality of the relevant evidence (sc. relevant to the question whether there is a sheep before her) that each woman has is *identical*; that evidence is sufficient to ground or support Lois's belief that there is a sheep before her; that *same* evidence is sufficient to ground or support Lana's belief that there is no sheep in the field at all. (Add, if you like, that each woman has considered the question whether there is a sheep before her carefully and at length, that each woman is cognitively unimpaired and that their deliberations on the 'sheep' in question were in each case wholly rational).

The second case entails that rational attention to evidence does not track the truth. The first does not.

12.2 Meta-Ontology: Reply to J. F. Göhner and L. Steinbrink

There is more to what the authors aptly call my "code of conduct" for those engaged in ontological disputes than "formalize, deduce and paraphrase", although these are certainly three important components of the code.² I will try to explain what I mean by this statement by considering an imaginary ontological dispute – a dispute between a nominalist and a platonic realist. Having considered and commented upon this imaginary dispute, I will attempt to generalize what I have said to all ontological disputes. I will close by commenting about several specific passages in the paper.

12.2.1 *The Imaginary Dispute*

My friend Norma is a nominalist: she says that there are no abstract, mathematical objects like numbers – and I say that there are indeed abstract mathematical objects, numbers among them. (Why is a dispute whether there are numbers "a dispute about ontology" and a dispute about whether there are odd perfect numbers not a

²As Captain Barbossa puts it in *Pirates of the Caribbean*: "The code is more what you'd call 'guidelines' than actual rules."

dispute about ontology? This is a very good question, which I shall not attempt to answer in the brief scope of this Reply.³)

I suggest to Norma that it would be advisable for us to take our discussion to the Ontology Room, and she agrees.

I will remark, parenthetically, that an ontological dispute or debate is not necessarily conducted in the Ontology Room. For such a dispute to be conducted in the Ontology Room is for the disputants to agree to follow a certain set of rules – a set of rules that were in large part, if not in their entirety, first clearly stated by Quine.⁴ And there are philosophers who refuse to play by any such rules. The late Ernest Gellner, for example, once wrote that, although he was a nominalist, he refused to conform to Quine’s ideas about what adherence to an ontological position like nominalism entailed:

The dreadful thing is, I haven’t even tried to be a serious, card-carrying nominalist. I have never tried to eliminate “quantification” over abstract objects from my discourse. I shamelessly “quantify over” abstractions *and* deny their existence! I do not try to put what I say into canonical notation, and do not care what the notation looks like if someone else does it for me, and do not feel in the very least bound by whatever ontic commitments such a translation may disclose. (Gellner 1979, 203)

This statement⁵ amounts to a refusal to enter the Ontology Room, a refusal to play by Quine’s rules – which a philosopher is, of course, perfectly free to refuse to do. And I, for my part, am perfectly free to refuse to discuss ontology with that philosopher. And I *would* refuse to discuss ontology with that philosopher – on the ground that I am very nearly certain (little is *entirely* certain in life) that the discussion would be a waste of my time.

Norma, however, is willing to play by the rules I regard as essential to a useful discussion of any ontological question. She accompanies me to the Ontology Room. Within its walls she says, “There are no numbers”. (In the Ontology Room, this sentence expresses the same proposition as the ‘canonical’ sentence “ $\sim\exists x (x \text{ is a number})$ ”.) What shall I say in reply? Shall I present her with an argument intended to prove the existence of numbers? I could do that – the Rules defining the Ontology Room do not say anything one way or the other about attempting to prove things. But that is not the way I work. What I would do is to try to convince her that some of the beliefs she brought with her into the Ontology Room, some things she believed before she ever opened any book on any philosophical topic, are inconsistent with her denial of the existence of numbers. That is to say, I would not present her with an argument whose conclusion is “There are numbers”, but rather an argument whose conclusion is “Some of the things you believe now and believed long

³But see note 8 in Sect. 12.2.2 of this Reply for a hint about how I would answer this question.

⁴For a partial statement of these rules as I would formulate them, see van Inwagen 2014a. For a debate that conforms perfectly to the rules of the Ontology Room, see Lewis and Lewis 1970.

⁵Here is my perhaps tendentious paraphrase of Gellner’s confession: “I shamelessly say things that logically imply the existence of abstract objects and deny the existence abstract objects! I do not feel in the least bound to accept the logical consequences of the things I say.”

before you became a philosopher logically entail the existence of numbers". I might for example present the following argument:

Look, you accept the propositions expressed by the following two sentences – the propositions they express here in the Ontology Room – , don't you?

- (1) Every ball – every solid object whose surface is a sphere – has a volume, which is equal to $\frac{4}{3} \pi$ times the cube of its radius.
- (2) There are balls.

And don't those two propositions jointly imply the proposition expressed by the sentence

- (3) $\exists x$ (the volume of $x = \frac{4}{3} \pi \times$ the cube of the radius of x)?

And doesn't the sentence

- (4) $\exists y \exists x$ (the volume of $x = y \times$ the cube of the radius of x)

follow from (3) by textbook logic? Now surely anything that satisfies " $\exists x$ (the volume of $x = y \times$ the cube of the radius of x)" must also satisfy " y is a number"? If so, then if the proposition expressed by (4) is true, something satisfies " y is a number". And, therefore, if the propositions you accepted before you entered the Ontology Room – propositions you accepted before you began your study of philosophy, and which your study of philosophy has given you no reason to reject – are true, there must be something that satisfies " y is a number". And, therefore, those propositions are inconsistent with nominalism. Unless you are willing to embrace a contradiction, therefore, you must *either* cease to accept certain of those propositions (e.g., that the volume of a ball is equal to $\frac{4}{3} \pi$ times the cube of its radius) *or* cease to be a nominalist.

And now the ball (whatever its volume may be) is, as I have put the matter elsewhere, in Norma's court. She may be able to return it successfully. Whether she indeed can is a question I will not enter into in this Reply. I turn instead to the question of what Norma, or someone in Norma's position, might say in the Ontology Room about *my* position, my platonic affirmation of the existence of numbers. One of the more interesting things a nominalist might say to me in the Ontology Room is this:

You say that there are numbers – that is, you endorse the proposition expressed by the canonical sentence " $\exists x$ (x is a number)". That is, you say that there are objects that satisfy the condition " x is a number". But what are those objects *like*? That is, what properties do they have? (Of course, I deny the existence of properties – I will explain later how I would ask that question without even apparently presupposing the existence of properties.⁶) Every real object – that is, every object: there are no unreal objects, as I am sure you will agree – has, for every property either that property or its negation. I do not think you can assign a coherent set of properties to numbers – to numbers in general or to any particular

⁶I wish my conversation with Norma had got to the point at which she presented that explanation. I'd very much like to know what it was.

number. What I want to see is your *theory* of numbers, your *account* of numbers, your statement of how they fit into reality as whole. I do not think you can give any such theory or account or statement. I do not think any of those things is *possible*.

And now the ball is in *my* court. It is my task to provide a descriptive (not an explanatory⁷) theory of numbers. I think I can return the ball, but whether I can is not a question to be answered within the scope of this Reply.

12.2.2 *A Schematic Generalization of the Imaginary Dispute About Nominalism*

There are, as I see matters, two parts to a dispute about whether there are *Fs*, where “*Fs*” represents some plural noun-phrase of the kind that figures in ontological disputes – “propositions”, “attributes”, “numbers”, “temporal parts”, “mereological fusions”, and so on.⁸ Let us imagine, in schematic form, the simplest dispute of that kind: there are two disputants, one of whom affirms the existence of *Fs* and one of whom denies the existence of *Fs*. Alfred, let us say, affirms the existence of *Fs* and Denise denies the existence of *Fs*.

There are two parts to the dispute because Alfred must do two things – at least if the dispute is to be a philosophically interesting dispute. First, he must show, by analysis of our pre-metaphysical discourse, that it commits us to the existence of *Fs* – that is, that our ‘everyday’ discourse, when properly understood, can be seen logically to imply that there are *Fs*. But doing this – if he *can* do it – will tell us very little about the *nature* of *Fs*. It will, as it were, show that an ‘*F*-role’ is inextricably involved in our everyday discourse, but it will not tell us much about the things that play the *F*-role – or about what properties things that *could* play the *F*-role would have to have.

Although I am now trying to present a very general schema that displays the form of ontological disputes, an illustration may be helpful in making this last point clear. Let us suppose that Alfred has demonstrated, or thinks he has, that our pre-metaphysical discourse logically implies the existence of *shadows* (two dimensional regions of relative darkness on the surfaces of solid objects that sometimes change their positions on those surfaces while retaining their identities) – that a ‘shadow

⁷I deny that explanatory theories have any place in metaphysics – a point I shall presently take up in the text. But not all theories are explanations. I have written a paper called “A Theory of Properties” (van Inwagen 2004), but the “theory” presented in that paper does not purport to explain anything – that is, it does not purport to explain how something-or-other can be the case. It does not, for example, purport to explain how it can be that two objects that are not identical in number are nevertheless identical in some respect – identical in color, identical in shape, and so on.

⁸Disputants in the Ontology Room may find themselves disputing about the existence, of, say, chairs – but only as ‘representatives’ of some much more inclusive category. A dispute about the existence of chairs that could not as easily have been a dispute about the existence of tables would not be of any ontological interest.

role' is inextricably involved in ordinary practical human discourse. His demonstration, even if it is successful, will leave unanswered many questions about the attributes of these shadows. (E.g., can one and the same shadow be cast by a tree in sunlight and, later, by that same tree in moonlight? If shadows are real objects, and if every real object has, for every property, either that property or its negation, then either a given shadow has the property "can be cast by two different sources of light" or it has the property "cannot be cast by two different sources of light.")

Alfred's second task, then, is to provide a descriptive theory of *Fs* – a theory that states what *Fs are* – and which, in virtue of having that property, enables us to answer philosophical questions about *Fs*. (But what does "a theory that states what *Fs are*" mean? That question cannot be answered in this brief Reply. I offer the theory of properties presented in the appropriately named "A Theory of Properties" (van Inwagen 2004) as a model for understanding what I mean by "stating what *Fs are*".)

Denise will certainly be involved in the first part of the dispute: it will be her job (her job as a defender of 'anti-Fism'⁹) to try to show that Alfred has failed to demonstrate that "an '*F*-role' is inextricably involved in our discourse". Or that is *one* thing she might do. Another possibility for her is this: she might concede that the *F*-role *is* inextricably involved in our discourse, but only in certain *parts* of our discourse, and declare her willingness to abandon the parts of our discourse that involve that role. And she may also have something to say to Alfred in the second part of the dispute: she may try to subject Alfred's theory of the nature of *Fs* to various dialectical pressures: she may try to show that the particular theory he presents is in some way defective. Or she may be more ambitious: she may, for example, try to show that the *idea* of 'an *F*' is incoherent – and hence that there cannot be a coherent descriptive theory of *Fs*. Or she may grant that the idea of an *F* is a coherent idea, but present arguments of some sort for the non-existence of *Fs*. And, of course, if there are no *Fs*, there can be a descriptive theory of *Fs* only in the sense in which there can be a correct descriptive theory of unicorns.

In short, what goes on in the Ontology Room is *philosophical argument* – the presentation of arguments for and against the positions under dispute. The purpose of the rules constitutive of the Ontology Room is summed up in the final sentence of "Being, Existence, and Ontological Commitment":

If these 'rules' are not followed, then – so say those of us who are adherents of Quine's meta-ontology – it is almost certain that many untoward consequences of the disputed positions will be obscured by imprecision and wishful thinking. (van Inwagen 2009b, 506)

⁹Of course, she might resign: she might, as a consequence of attending to Alfred's arguments for the conclusion that certain of her beliefs are inconsistent with anti-Fism, cease to be an anti-First. I will assume that this is not the course she chooses: she is not convinced by Alfred's arguments and undertakes to answer them.

12.2.3 *Comments on Some Specific Passages in the Paper*

I will now comment on several specific passages of the paper.

Passage 1

Recall that van Inwagen's proposed account of meta-ontology asks us to provide a formal paraphrase for any ordinary language sentence the existential implications of which we are not willing to accept. (Sect. 3.2.1)

I am in fact willing to accept the existential implications of most sentences of natural language that anyone has ever actually used as a vehicle of assertion. And the existential implications of 'ordinary' assertions that I regard as false, I regard as false for reasons that have nothing to do with philosophy (for example, certain of the existential implications of "The pyramids were constructed by extraterrestrial beings"). In "Inside and Outside the Ontology Room" (van Inwagen 2014a), I imagined the following exchange (an exchange, of course, that takes place outside the Ontology Room).

"You and I may be brothers, but no two people could be less alike. I have devoted my life to working for peace and justice, and your only goal in life is to get rich selling furniture."

"What can I say? I deal in reality and you deal in dreams. Chairs exist. Peace and justice don't and never will." (van Inwagen 2014a, 5)

I accept all the existential implications of what the second speaker asserted when he said, "Chairs exist". But if the same speaker said, "Chairs exist" *inside* the Ontology Room, I would dispute his assertion. Spoken inside the Ontology Room (I contend) that sentence expresses a proposition that is true only if something satisfies the open sentence "*x* is a visible tangible composite object that is not alive" (and, of course, it is my position that nothing satisfies this sentence). The proposition the second speaker *actually* expressed, however, does not imply that anything satisfies that open sentence. That proposition has, or so I say, precisely the same existential implications as the proposition expressed by "There are things that are arranged chairwise" – and I accept the existential implications of that proposition. (But it is not my view that these propositions are the same. If a French royalist, in a debate with a royalist who holds a different view about the legitimate claimant to the French throne, says, perhaps alluding to well-known physical features of the rival claimants, "*Le roi actuel de la France est chauve*," that assertion – so at any rate I maintain – has the same existential implications as those of the proposition expressed by "Something is male and now legitimately reigns over France and everything that is male and now legitimately reigns over France is identical with that thing and it is bald". I should not want to say, however, that these two sentences expressed the same proposition.)

Passage 2

This is odd: although he warns us that the surface structure of ordinary language may lead us astray ontologically, van Inwagen still defends a meta-ontological view that relies heavily on our everyday utterances. (Sect. 3.2.1)

Well, as J. L. Austin said: “Certainly, then, ordinary language is *not* the last word: in principle it can everywhere be supplemented and improved upon and superseded. Only remember, it *is* the *first* word.” (Austin 1961, 133) My hard-headed, cynical businessman, speaking outside the Ontology Room, said, “Chairs exist”. What he asserted by saying that (what *I* say he asserted by saying that) is not very interesting. It is a truism – something no one would dispute.¹⁰ What *is* interesting (I say) is what he would say inside the Ontology Room – for, in addition to being a businessman, we may imagine that he is also a metaphysician and is willing to enter the Ontology Room. Suppose that, like many metaphysicians, he is willing to say “Chairs exist” not only outside but inside the Ontology Room – that is, that he is willing to say something that implies that something satisfies the open sentence “*x* is a visible tangible composite object that is not alive” (which was not, I contend, an implication of the proposition expressed by his ‘outside’ utterance of “Chairs exist”). I do not think that the concept “ordinary language” is precise enough for there to be a determinate answer to the question whether speakers in the Ontology Room are speaking “ordinary language”. What I *am* willing to say is that discourse in the Ontology Room is not ordinary *discourse* – for discourse inside the Ontology Room is subject to a set of rules that do not govern discourse outside the Ontology Room. And what I am willing to say about whether “ordinary language is a reliable guide to ontology” could be put in the form of the following question and answer:

- Q: Is the truth-value of the proposition expressed by a sentence uttered outside the Ontology Room a reliable guide to the truth-value of the proposition expressed by that same sentence when it is uttered inside the Ontology Room?
 A: No.

Nevertheless, ordinary discourse is what we must start with. All the rules that govern the extraordinary discourse that takes place in the Ontology Room are stated in everyday language. All the technical terms used in the Ontology Room (“quantifier”, “scope”, “variable”, “open sentence”, “satisfy”, “object”, and so on) are defined or explained in everyday terms. After all, we must create the Ontology Room before we can enter it, and the only linguistic resources that we can utilize in its creation, the only ones available to us, are those of everyday speech.

Passage 3

On van Inwagen’s account, the burden of proof is with the nominalist, who must demonstrate that a given ordinary language sentence does *not* commit him or her to the existence of some (sort of) entity. Concerning the nominalist’s denying the existence of certain mathematical objects, van Inwagen clearly states that “the ball is in their court” [...]. Realism (concerning any entity that can be named), on this view, is introduced as the default position; as such, it is privileged over nominalism. Unless the nominalist proves otherwise (e.g.,

¹⁰Some metaphysicians – Trenton Merricks, for example – will deny that when the businessman says, “Chairs exist” in the circumstances I have imagined he says something true. But these metaphysicians and I differ about the proposition he expresses when he speaks that sentence. They say it expresses a certain proposition *p* and I say it expresses a certain proposition *q* (where *p* entails *q*, but not vice versa). They and I agree that *p* is false. They and I agree that *q* is true.

that all ordinary language sentences about sakes can be paraphrased in a way that avoids a commitment to sakes), what he or she says is exactly what he or she believes. Prima facie, this strikes us as odd, as we would suppose the burden of proof to be with whoever wants to claim the existence of something [...]. (Sect. 3.2.1)

I am never sure what the legal term “the burden of proof” means when philosophers use it. Insofar as I can make any sense of it, I would say that the burden of proof falls equally on every philosopher, or at least every philosopher who makes controversial philosophical assertions he or she wishes other philosophers to accept. If, for example, I say that there are abstract objects, and if I want other philosophers to accept that assertion, I should be prepared to offer an argument for the existence of abstract objects – or, at any rate, an argument for the conclusion that various things they believe entail the existence of abstract objects. If Norma says that there are no abstract objects, and if she wants other philosophers to accept that assertion, she should be prepared to offer an argument whose purpose is to convince them that they should not believe that there are abstract objects. I see no reason to suppose that “the burden of proof [is] with whoever wants to claim the existence of something”. You might, of course, say that if practically every rational person believes that there are no *F*s, then the burden of proof falls on those who maintain that *F*s exist. But if you say that, by the same token, you should say that if practically every rational person believes that there are *F*s, then the burden of proof falls on those who maintain that *F*s do not exist. (Either the existence or non-existence of *F*s might be the ‘default’ position. It would depend on what “the *F*s” were.) And what if there are plenty of rational people who believe that *F*s exist and also plenty of rational people who believe that *F*s do not exist? Then, surely, “the burden of proof”, whatever precisely that is, falls equally on both parties. And that is how things stand with realists and nominalists: both those schools are well stocked with perfectly rational people.

In the imaginary dispute in Sect. 12.2.1, I presented an argument for the conclusion that many of the beliefs that Norma brought to philosophy entailed the existence of numbers. Having done that, I said that the ball was in her court. I meant no more by that than, “Here is an argument for the conclusion that that either you should abandon those beliefs or abandon nominalism. What is your response to it?” And I imagined Norma’s demanding that I provide a coherent account of the nature of numbers – and said that, when she has made that demand, the ball was in *my* court.

I am happy to grant that there are ‘default’ positions in philosophy. I take the proposition that conscious thinking beings exist to be a default position in ontology. In my work on free will, I have taken the position that the falsity of “Nothing that has ever happened is anyone’s fault” is a default position. But I do not think that either the nominalists or the realists have any right to regard their position as the default position in the matter of the existence of abstract objects.

Passage 4

The propositions expressed through ordinary language sentences do commit us to the existence of entities (in the broadest sense of the word), but these entities do not necessarily correspond directly to the expressions’ supposed referents. This, we suggest, is the greater value of the proposed method to van Inwagen: to establish coherence between metaphysical

discourse and everyday talk. Although this aim may well be worthy of pursuit, it relies on certain presuppositions that have to be established independently. Among these presuppositions is the claim that we are capable of telling our ontological commitments apart from the byproducts of language, and that this is possible independently of the three-step methodology as introduced by van Inwagen. (Sect. 3.2.1)

I would put the “presupposition” this way: Following the rules in force in the Ontology Room will ensure that one’s ontological commitments are not “byproducts of language”. Consider, for example, the dispute between “Argle” and “Bargle” about the existence of holes (Lewis and Lewis 1970) – a paradigm case of an ontological dispute in which those rules are strictly observed. Would anyone suppose that Argle’s “commitment” to the existence of things he calls “holes” was a byproduct of language? I reject Argle’s ontology root and branch, but that it is a byproduct of language is one of the few negative things that might be said about it that *I* would not say.

Passage 5

What do we expect of an ontological theory for it to be successful? If, according to van Inwagen, explanation is in no way the standard to be applied, what is to replace it as a measure for the achievements of a proposed ontology? (Sect. 3.2.2)

Well ... that it is internally consistent; that it contradicts nothing we know to be true on non-metaphysical grounds; that either it is consistent with all the default positions in ontology (e.g., that conscious thinking beings exist) or presents reasonable grounds for supposing that the default positions it denies do not deserve that status; that it incorporates convincing replies to all the arguments against it ... These are measures of the *achievements* of a proposed ontology. (Whether a given ontological theory *has* achieved these things will, of course, be controversial. But this is philosophy: *every* substantive, positive philosophical thesis is controversial.) I am not sure what it means to say that an ontological (or, more generally, philosophical) theory is *successful*. I can think of various ways to define “successful ontological theory”. They would be adaptations of the various definitions of “successful philosophical *argument*” that I considered in Lecture 3 of *The Problem of Evil* (van Inwagen 2006). Unfortunately, these definitions would one and all have the consequence that no substantive ontological theory¹¹ was a success – and it seems to me, therefore, that the question whether a given philosophical theory (a category that of course includes all ontological theories) is a “success” is an idle question.

I will remark, finally, that the authors do not tell their readers why I think that that it is not the business of metaphysical theory to explain things. I think the reason is worth a brief statement. And it is: having examined the questions that explanatory metaphysical theories purport to provide answers to, I have reached the conclusion that it makes no sense to suppose that these questions have answers. Consider, for example, the ‘question’ I referred to in note 7 above: How can it be that two objects that are not identical in number are nevertheless identical in some respect—identical

¹¹“Something exists” might be a ‘successful’ ontological theory. But if that is a ‘theory’ at all, it is certainly not a substantive one.

in color, identical in shape, and so on? Something I have said elsewhere (van Inwagen 2011, 396) applies to this ‘question’: “[N]o meaningful statement among all possible meaningful statements counts as an answer [to it]”. I put the word “question” in scare-quotes here, because, as Wittgenstein has said:

Zu einer Antwort, die man nicht aussprechen kann, kann man auch die Frage nicht aussprechen. *Das Rätsel* gibt es nicht. Wenn sich eine Frage überhaupt stellen läßt, so *kann* sie auch beantwortet werden.

When the answer cannot be put into words, neither can the question be put into words. *The riddle* does not exist. If a question can be framed at all, it is also *possible* to answer it. (Wittgenstein 1921, Tractatus 6.5, transl. Pears/McGuinness)

12.3 Physical Composition: Reply to J. Husmann and P. M. Näger

12.3.1 Counterexamples as Arguments

The authors contend that I give no arguments for the falsity of the following answers to the Special Composition Question: Contact; Fastening; Cohesion; Fusion. (Let us call these principles collectively Proposed Bonding Answers, or simply Answers.) They do not of course deny that I present what I consider counterexamples to these four Answers – for they quote them in full. Why are these counterexamples not arguments for the falsity of these Answers? Does a philosopher’s presenting what he or she considers a counterexample to a general principle not count as presenting an argument against it? The objection appears to be this: Each proposed counterexample is an imaginary case in which the Answer under consideration implies that composition occurs – and in each case I contend that composition does *not* occur; but I give no arguments to show that composition does not occur in these cases.

Let us consider an analogy. A philosopher proposes the following thesis: Reality contains no case of two numerically distinct objects – that is to say, for the thesis that numerical distinctness is an illusion. A second philosopher responds by saying (with two appropriate gestures), “Here is a hand ... and here is another hand.” Must that second philosopher *also* be required to present an argument for the proposition, “I have just displayed a state of affairs that contains two numerically distinct objects”? Well, someone might *want* the second philosopher – let us call him “George” – to do that. Suppose, however, that George confesses that he is unable to produce any argument for that conclusion whose premises are more evident than the conclusion itself. Still, the counterexample George has proposed still counts as an argument against “Reality contains no case of two numerically distinct objects”. Perhaps not every philosopher will be convinced by this argument. It is no secret that an argument that some philosophers find absolutely convincing other philosophers will say “is guilty of a subtle logical fallacy” or “assumes the very point at issue” or “rests on a premise that there is simply no reason to accept” or “employs

the same term in different senses at different points in the argument” or ... (fill in the remainder of the list yourself). All I can say is what I have said: it seems to me that the counterexamples I have proposed to the four Proposed Bonding Answers are decisive – they *show* that those Answers are wrong. If the authors do not find them decisive, or even remotely plausible, I can only say that that is how things go in philosophy.

12.3.2 *Fastening*

Consider “two rings that are forged in an entangled way”:

This example also shows that van Inwagen’s partition of bonding into fastening, cohesion and fusion is incomplete; for the entangled rings do not fall into any of his categories (though they are clearly bound to another). (Sect. 4.7.1)

This was my definition of “fastening”:

Suppose that two objects are in contact and suppose that they are so arranged that, among all the many sequences in which forces of arbitrary directions and magnitudes might be applied to either or both of them, *at most only a few* would be capable of separating them without breaking or permanently deforming or otherwise damaging either of them. Then let us say that these two objects are *fastened to each other* or simply *fastened*. (van Inwagen 1990, 56)

The two-rings example shows that this definition is flawed – but the flaw is a minor one and easily mended; to accomplish the needed repair, simply replace “in contact” with “in close proximity”. Then the two rings are fastened. (There is no way of separating them – of causing them not to be “in close proximity” – without damaging either of them, and there are therefore at most only a few ways of separating them without damaging them: there are, at present, at most only a few kings of France.) And there is, therefore, (once this minor revision has been made) no reason to suppose that my partition of bonding is incomplete.

Consider a woman who has made a circle, using her thumb and middle finger. Consider a second woman who has done the same. Let them make their finger-circles in such a way that the two circles are ‘entwined’ (perhaps this mutual action is the secret sign of recognition of some esoteric sisterhood to which they both belong). Let their fingers, at that moment, become paralyzed. Then, at the moment the paralysis occurs, the two women become fastened to one another. But, or so *I* contend, nothing such that

- (i) both women are parts of it

and

- (ii) each of its parts overlaps at least one of the women

comes into existence at the moment the paralysis occurs. (You might believe that there was a thing with those mereological properties, but – I contend – if you believe that this object existed after the onset of the paralysis, you should believe that it

existed before the onset of the paralysis as well.) The paralysis of their fingers, or so I maintain, “has not added to the furniture of earth; it has merely diminished its capacity to be rearranged” (van Inwagen 1990, 58).

12.3.3 *Science Is Mereologically Neutral*

According to a standard contemporary scientific view, however, there are organs that are proper parts of complex organisms and have cells as their proper parts; (bio-) molecules are parts of cells; atoms are parts of (bio-)molecules; protons, neutrons and electrons are parts of atoms; these electrons are simples according to the actual standard model of particle physics; protons and neutrons, however, have up- and down-quarks as parts, which, finally, are simples as well. In sum, what science suggests is that there are many more inanimate objects in between simples and organisms than van Inwagen accepts, and that these objects form a *nested hierarchy of parts and wholes* from quarks and electrons up to macroscopic organisms. (Sect. 4.3)

To my ear, there is a difference between “Theory X suggests that *p*” and ‘According to Theory X, *p*’. I do not know what to make of assertions like “Science suggests that there are inanimate objects that are not simples”. I expect that what science *suggests* is in the eye (or ear) of the beholder. To take a particular case, I do not know what to make of the assertion that the Standard Theory ‘suggests’ that there are protons. I *do* know what to make of the following two assertions, however – which are presumably stronger than the “suggests that” assertion (whatever it may mean) *and* are equivalent to each other – are two ways of saying the same thing:

- (1) The Standard Theory logically implies that there are protons.
- (2) According to the Standard Theory, there are protons.¹²

I understand these statements well enough to be of the firm opinion that they are both false – or at any rate false if “there is” is understood in its “strict and philosophical sense” (if the two statements are indeed equivalent they must, of course, have the same truth-value). Consider these two statements:

- (3) There are *xs* such that those *xs* are three in number and one of those *xs* is a down-quark and two of those *xs* are up-quarks and those *xs* are arranged (or if you like, “bound”) protonically.¹³
- (4) There are *xs* such that those *xs* are three in number and one of those *xs* is a down-quark and two of those *xs* are up-quarks and those *xs* are arranged protonically; *and* any three quarks arranged protonically have a mereological sum.

¹²An extreme pedant might want to add an environmental qualification – something like “unless the temperature everywhere in the cosmos is greater than 2×10^{12} degrees K” – to both these statements.

¹³For the meaning of “the *xs* are arranged protonically” consult the Standard Theory. Its textbook formulations will not, of course, contain that *phrase*, but they will precisely describe the mutual causal relations that I mean the phrase to express.

It is my firm opinion that, although both the Standard Theory and statement (4) imply statement (3), the Standard Theory does *not* imply (4) – nor does it imply the falsity of (4). In other words, the Standard Theory does not ‘care’ whether (4) is true. The Standard Theory is mereologically *neutral*. The Standard Theory is mereologically *blind*. And here is the reason why the Standard Theory does not care whether (4) is true: If (4) is true – that is, if there are (as existential propositions are understood in the Ontology Room¹⁴) protons, physicists are interested only in such matters as their mass, their electrical charge, and the various fields of force they generate (gravitational, electromagnetic, weak- and strong-nuclear). But the rest mass of a proton is nothing other than the sum of the relativistic masses of its constituent quarks¹⁵ (and so, mutatis mutandis, for its charge), and the vector fields for any given kind of force or interaction generated by a proton is simply the vector sum of the fields generated by each of its constituent quarks (and such bosons as may be transmitting interactions among them). Thus, all physical interactions in the cosmos ‘look’ the same whether or not three quarks arranged or bound protonically have a mereological sum.

This point may be easier to grasp if we consider a physical theory that is simpler than the Standard Theory, classical gravitational mechanics. Consider the planet Jupiter. (And adopt the following idealization: Jupiter is a mathematically perfect homogeneous ball of uniform density – a ball with a radius of about 6.9911×10^7 m and a mass of about 1.898×10^{27} kg.) An important theorem of classical gravitational mechanics implies the statement (the time variable is suppressed):

Consider Jupiter’s gravitational field; consider the gravitational field that *would* be generated by a punctiform object whose mass was equal to Jupiter’s mass and which was located where Jupiter’s center of mass is located (or located where the geometrical center of Jupiter is located – these are the same under the conditions of our idealization); *outside* the surface of Jupiter, these two fields are identical (have the same value – an acceleration, so-and-so many meters/second/second, in the direction of the center of Jupiter – at every point in space).

This is an extremely useful result: if we are concerned with the gravitational force Jupiter is exerting on some object outside its surface – the moon Ganymede, let us say –, our calculations do not have to take account of the fact that Jupiter is an immense ball, a ball whose radius is a significant fraction of the distance from its center to Ganymede. We can, rather, for the purposes of our calculation, treat Jupiter as a punctiform object with a mass of 1.898×10^{27} kg located at the point in space occupied by the center of the ‘actual’ Jupiter.

And how is this extremely useful result proved? Simplifying somewhat, in the following manner. Imagine that Jupiter is a ball-shaped cloud of uniformly distributed point-sized particles of equal mass. Suppose, for example, that there are a billion particles in the cloud – in which case each of them will have a mass of 1.898×10^{27} kg divided by 1 billion kilograms, i.e., 1.898×10^{18} kg, and each will

¹⁴ See Sect. 12.2 for the meaning of “the Ontology Room”.

¹⁵ Plus the mass equivalent of their binding energy – which is *enormous*.

be at some determinate position in space. Now describe the gravitational field generated by each of these billion particles and calculate the vector sum of these billion individual fields. Call this field F_1 . Now increase the number of particles and do the same thing again, obtaining the field F_2 . Do this repeatedly. As the number of particles in the cloud increases without limit, it turns out that the sequence of fields will converge to a “limit field,” F_{limit} . It *also* turns out that F_{limit} is identical with the field generated by a single punctiform particle with a mass of 1.898×10^{27} kg located at the center of Jupiter – identical, that is, at all points more than 6.9911×10^7 m from the center of Jupiter (i.e. all points outside the surface of Jupiter). This is the general form of the proof (with all the hard parts left out).

Now consider a point in the sequence where the number of particles is *very* large – say, 10^{50} . The sum of the fields generated by the individual particles in this case – call it F – will differ from F_{limit} by values that lie far below the possibility of detection by any measurement. Now suppose that Jupiter actually *is* composed of 10^{50} uniformly distributed particles. Call Jupiter’s gravitational field F_J . Note that the intuition behind the way the above result was obtained implies that F and F_J are *identical* – are one and the same gravitational field.¹⁶ In other words, classical gravitational mechanics is mereologically neutral, mereologically blind. Classical gravitational mechanics does not *care* whether those 10^{50} particles have a mereological sum. Classical gravitational mechanics does not *care* whether there is a vast ball called “Jupiter” or only a lot of ‘ballwise-arranged’ particles (‘ballwise’ is the only stable spatial arrangement possible for particles in their situation¹⁷) that have fallen into their mutual gravitational well. There is simply no physical difference between those two (epistemic) possibilities. Metaphysics cares whether Jupiter exists ‘in the strict and philosophical sense’. Physics does not.¹⁸

12.3.4 *Against Composition by Bonding*

All the supposed counterexamples to the various “bonding” answers to the Special Composition Question (SCQ) that I presented in §§ 3, 6, and 7 of *Material Beings* (van Inwagen 1990, 33–37 and 56–71), if they are indeed counterexamples to the those bonding answers, are also counterexamples to the authors’ proposed answer. I now present a counterexample specially tailored to their answer.

¹⁶Both outside *and* inside the surface of Jupiter. If it were possible to drill a hole through Jupiter from pole to pole, and if you dropped, say, an iron cannonball into that hole, the gravitational force acting on the ball as it fell, passed the center of Jupiter, and rose toward the opposite pole, would at every moment be the sum of the gravitational force each of the 10^{50} particles was then exerting on it.

¹⁷I neglect the effects on Jupiter’s shape of its rotation about its polar axis.

¹⁸For more on this topic, see Chapter 7 (“Is There Scientific Evidence for Composition?”) of Brenner 2016.

Suppose I am wearing a space-suit. Let us refer to the sum of me and the space-suit and anything fastened to the space-suit as ‘I⁺’ or ‘me⁺’ (as grammar requires; I trust the meaning of the possessive pronoun ‘my⁺’ is clear). I grant for the sake of argument that I⁺ exist. Suppose that I⁺ am moving rapidly on a vector that, if it were not for gravitational forces, would result in my⁺ moving in a straight line that would pass close to some small (by astronomical standards) but very massive object – let us say a neutron star. The gravitational field of the neutron star, however, does not allow my⁺ path to be a straight line: it bends my⁺ path into a hyperbola with my⁺ and the star’s common center of mass (for all practical purposes the star’s center of mass) at its focus – for my⁺ kinetic energy exceeds my⁺ gravitational potential energy with respect to the star. (My⁺ kinetic energy at any moment *t* is one half my⁺ mass times the square of my⁺ velocity at *t*. My⁺ gravitational potential energy at *t* is the amount of energy it would take to raise a body with my⁺ mass from the center of the star, working against the gravitational pull of the star – if this were somehow possible – to my⁺ position at *t*. Or we could say my⁺ potential energy at *t* was equal to the kinetic energy that a stationary body with my⁺ mass that occupied my position at *t* would have after it fell to the center of the star – assuming, again, that that was really a possible thing for a body to do.)

Now here is something I did not mention. Clipped to the outside of the space-suit, there is a tool of some sort – let us say, a hammer. Call the sum of the space-suit and the hammer Suit⁺. The following objects are in play in our story – again, I will assume for the sake of argument that they all exist:

I + the space-suit; the hammer, Suit⁺; I⁺; the neutron star.

In the circumstances described, the authors’ answer to the SCQ yields the following result: the following fusions of objects in the above list (and no others) exist.

I + the space-suit

The space-suit + the hammer (= Suit⁺)¹⁹

I + the space-suit + the hammer (= I⁺).

Now suppose that for some reason I do the following: I unclip the hammer from the space-suit and I throw it *very hard* in the direction in which the space-suit and I are traveling. (This has the immediate result that I⁺ and Suit⁺ cease to exist, but this fact will not figure in my argument.) Thus, the magnitude of the momentum of me + the space suit in the direction of travel is decreased and the magnitude of the momentum of the hammer in that direction is increased – and by the same amount. This implies that the velocity of me + the space-suit decreases. The mass of me + the space-suit, moreover, is less than the mass of me + the space-suit + the hammer – the difference of course, being equal to the mass of the hammer. Therefore,

¹⁹ I think this is correct – but perhaps the “no-extension” clause that figures in the criterion of physical composition (Sect. 4.6) implies that I + the space suit and the space-suit + the hammer do not exist. Whether these fusions exist, however, is not relevant to my argument. I will assume the existence of I + the space-suit in the sequel (for ease of exposition), but this assumption could, if necessary, be eliminated by paraphrase.

after I throw the hammer, the kinetic energy of me + the space-suit will be less than the kinetic energy of me + the space-suit + the hammer before I threw the hammer. If only I throw the hammer *hard* enough, I shall be able to reduce the kinetic energy of me + the space-suit till it is *less* than that object's gravitational potential energy with respect to the star.²⁰ Suppose, then, that I *do* throw it 'hard enough' – but not *too* hard. (If I threw it *really* hard, I could reduce my speed sufficiently to bring me + the space-suit to rest with respect to the star – in which case I + the space-suit would fall in a straight line toward the center of the star; if I threw it less hard than that, but still 'too hard,' I + the space suit would then 'spiral in' till it struck the surface of the star – or till various bits of it individually struck the surface of the star: it would probably be pulled apart by tidal forces before it reached the surface.) Suppose, that is, that I threw the hammer with a velocity that had this consequence: I + the space suit's trajectory was changed from a hyperbola to an ellipse (with the star at one focus) – that is to say, with a velocity that put me + the space-suit into orbit around the star.

When I + the space-suit is in orbit around the star, the former is gravitationally bound to the latter. According to the authors' proposal, then, at the moment the hammer leaves my hand, a composite object, namely,

(I + the space-suit) + the neutron star.

comes into existence.

It is a basic conviction of mine that one cannot bring something into existence simply by *throwing* something – and that the authors' answer to the SCQ is therefore wrong. I do not, of course, claim to have presented a knock-down argument for the falsity of their theses about composition. It is very rare for there to be knock-down arguments for the falsity of a substantive philosophical thesis.²¹ All I have done, in fact, is to deduce a proposition that seems to me to be obviously false from their answer to the SCQ – and they, I am fairly sure, will say that that not only do they not find that proposition obviously false, they find it obviously true. Well, that is how things go in philosophy. If I have done anything of interest in this Reply, I think it is this: in Sect. 12.3.3, I have called into question the point of view concerning the ontological implications of science that is presented in the quotation with which that section begins.

²⁰At any given point in space, the gravitational potential energy of me + the space-suit with respect to the star is less than that of me + the space-suit + the hammer (since the mass of former is less than the mass of the latter). But the decrease in potential energy depends only on the mass of the hammer, whereas the decrease in kinetic energy depends both on the mass of the hammer and the velocity with which it is thrown. If it is thrown fast enough, the decrease in kinetic energy can greatly exceed the decrease in potential energy.

²¹"Gödel and Gettier may have done it." (Lewis 1983, p. x)

12.4 Artifacts: Reply to A. Bahr, C. Fischer, T. Kater and N. Kleinschmidt

12.4.1 *Non-existence of Artifacts and Common Sense*

Our everyday view of the world has it that we are surrounded by a vast number of objects such as tables, chairs, cars, books, and smartphones – objects that philosophers frequently refer to with the term “artifact”. To hold “that there are no tables or chairs or any other visible objects except living organisms” [...] radically clashes with our common sense convictions regarding our environment. (Sect. 5.1)

For reasons that I went into at some length on pages 103 and 106 of *Material Beings* (van Inwagen 1990), I deny that “common sense” has any metaphysical implications. Common sense tells one not to venture casually into dangerous neighborhoods late at night and to make a yearly visit to one’s dentist. It does not tell us whether any of the propositions debated by metaphysicians is true or false.²² There is, however, such a thing as Universal Belief: “that body of propositions that has been accepted by every human being who has ever lived, bar a few imbeciles and madmen.” (p. 103) Here is what I regard as a *useful* philosophical question:

Does the thesis that there are no artifacts contradict any of the constituent propositions of Universal Belief?

My answer is No (van Inwagen 1990, Preface and §10).²³ I will not recapitulate the arguments presented in *Material Beings* and “Inside and Outside the Ontology Room” (van Inwagen 2014a) for this conclusion here, but here is a brief argument for the same conclusion:

If the thesis that there are no artifacts contradicted any of the constituent propositions of Universal Belief, it would contradict a very high proportion of them. If it contradicted a high proportion of them, then most of the assertions speakers made while engaged in “the ordinary business of life” would be false. But it is not the case that most of the assertions speakers make while engaged in the ordinary business of life are false. (Consider, for example, the imaginary exchange between two brothers – quoted in Sect. 12.2.3 from van Inwagen 2014a). When the cynical brother said, “Chairs exist,” what he said was, I contend, *true*.) Therefore, the thesis that there are no artifacts does not contradict any of the constituent propositions of Universal Belief.

²² My metaphysic of material objects is therefore not what Sir Peter Strawson called a revisionary metaphysic. (Is it therefore a descriptive metaphysic? I would not call it that, either. I consider Strawson’s taxonomy of metaphysical positions to be incomplete.)

²³ See also van Inwagen 2014a.

12.4.2 *The Dialectical Structure of Material Beings*

Do I present arguments for the non-existence of artifacts in *Material Beings*? That is a question that has to be approached with some care. The position defended in that book is this: My proposed Answer (I will use “Answer” as an abbreviation of “Answer to the Special Composition Question”) is the only known Answer that is consistent with the following ten “constraints” (see the Preface) and does not entail obviously false theses about the existence of composite objects (e.g., that if a cruel child glues two hamsters together, this act brings into existence an object of which the two unfortunate creatures are parts):

1. The classical or ‘absolute’ view of the identity relation (as opposed to Geachian ‘relative identity’) is correct.
2. Material things persist through time (every material thing *literally* and ‘in the strict and philosophical sense’ exists at more than one time), and endurantism (as opposed to perdurantism) is the correct metaphysic of persistence.
3. Standard, textbook logic is correct – at least in cases that do not involve vagueness.
4. Things have objective modal properties *de re*; everything exists in more than one possible world, and, although some of the properties of any given object will vary from world to world, there are, for each object, properties that object has in every world in which it exists – and, therefore, Lewis-style “counterpart theory” (Lewis 1986) provides an incorrect account of such assertions as “Descartes could have lived longer than he actually did”).
5. Matter is ultimately particulate.
6. Two objects cannot have exactly the same proper parts at a given moment.
7. There can be no thought or sensation unless mental predicates (“is in pain”, “is thinking of Vienna” ...) apply to individual things; there can be no sensation or thought unless individual things feel and think.
8. Thought and sensation *are* real features of the world, and the individual things that feel and think (or some of them: human beings, the referents of names like “René Descartes” and “Catherine the Great”) are living organisms; they are neither immaterial substances nor proper parts of living organisms (such as brains or cerebral hemispheres).
9. What there is is never a matter of stipulation or convention.
10. Whether certain objects compose some larger object depends only on the spatial and causal relations in which they stand to one another.

Now let us refine our vocabulary for talking about artifacts. Consider the example of the intelligent snake that is woven into a hammock (van Inwagen 1990, 126–127). Is the hammock so produced an artifact? Let us say that if it is, it is not a *normal* artifact. Normal artifacts are, or are if they exist, such things as tables, chairs, books, houses, computers, marble sculptures, wine glasses, sandwiches,

cigars, and aircraft carriers. That is, they are entirely composed of non-living material – although, as in the case of the sandwiches, the cigars, and at least some of the books, they may be wholly or almost wholly composed of once-living organic materials.

The following two statements are conceptual truths:

- (i) All normal artifacts are visible, tangible, non-living composite objects.
- (ii) The Proposed Answer to the Special Composition Question implies that there are no visible, tangible, non-living composite objects.

If the position defended in *Material Beings* is correct, therefore, every known answer to the Special Composition Question that is consistent with the ten “constraints” (and which does not entail such implausible existential-mereological theses as that gluing two animals together causes an object that is the fusion of the two animals and the particles that make up the layer of glue that binds them to come into existence at that moment) implies that there are no normal artifacts.

The arguments of *Material Beings* can therefore certainly be said to *support* or *favor* the thesis that there are no artifacts. But is there anywhere in *Material Beings* an argument whose conclusion is “There are no normal artifacts”? If there is, it would have to be something like this:

The Proposed Answer is correct.

If the Proposed Answer is correct, there are no visible, tangible, non-living composite objects.

If there are normal artifacts, there are visible, tangible, non-living composite objects.

Therefore, there are no normal artifacts.

But “The Proposed Answer is correct” is not one of the theses defended in *Material Beings*. (Well, perhaps it is in some loose sense of “defended”. But I have never defended it in the sense of presenting arguments whose conclusion is “The Proposed Answer is Correct”.) You can accept all the arguments of *Material Beings* without accepting that thesis. Accepting all the arguments of the book commits you only to the conclusion that the Proposed Answer is the only *known* Answer that satisfies certain conditions (the ones I have laid out above) – and that conclusion, of course, does not logically imply that the Proposed Answer is correct. (Perhaps the correct Answer is inconsistent with the Proposed Answer and no philosopher has ever thought of it. Or perhaps at least one of the propositions (1)–(10) is false. Or perhaps one *can* bring composites into existence by gluing animals together.)

Still, I do accept the proposition that the Proposed Answer is correct. And I do, therefore, *accept* the above argument for the non-existence of normal artifacts, even if I have never *offered* or *proposed* it. And I maintain that it is the *only* argument for the non-existence of normal artifacts that I accept. I maintain further that I have never offered or proposed any *other* argument for the non-existence of normal artifacts.

12.4.3 Clarifying the Conclusions of the “Arguments Against Artifacts”

The authors consider three arguments presented in §13 (“Artifacts”) of *Material Beings*, which they classify as “arguments for the non-existence of artifacts”:

- The Gollyswoggle Argument
- The Living Hammock Argument.
- The House of the Wise Pig Argument.

But these arguments are *not* arguments for the non-existence of artifacts. The first two of these arguments (van Inwagen 1990, 126–127) were directed against the following thesis (and similar theses):

When a sculptor kneads a lump of clay into a statue, a statue comes into existence. But (normally, at least) in such a case nothing goes out of existence. That is to say, the lump of clay continues to exist. When the statue comes into existence, therefore, it is spatially conterminous with the lump – but numerically distinct (by Leibniz’s Law: it has different historical and modal properties from those of the lump).²⁴

The conclusions of the Gollyswoggle Argument and the Living Hammock Argument are perfectly consistent with the existence of clay statues – and hammocks, living or not. Their conclusion is that the statue and the lump, *if* they ‘both’ exist, are numerically identical. That is to say, the Gollyswoggle Argument is a defense of the thesis that – if there are both lumps of clay and clay statues at all – “clay statue” is a *phase sortal*, a phrase that applies to lumps of clay only when they are of certain shapes (and, no doubt, only when they have been molded into the shape they have by an artist who had the prior intention of molding them into approximately that shape). Thus, the person who is idly working a piece of clay with his fingers is *not* “causing the generation and corruption of the members of a compact series of objects of infinitesimal duration” (van Inwagen 1990, 126) – for he is not causing the generation or corruption of anything: If there is anything in his hand at all during this episode, it is an enduring lump of clay, an object whose shape changes but whose identity does not. Or, to use an example of the authors, if one were to turn a small snake into a living bracelet (by persuading it to imitate the Worm Ouroboros), then, if one continuously changed the shape of the living bracelet by pressing it with one’s fingers – perhaps changing its shape gradually from ‘perfect circle’ to ‘ellipse of high eccentricity’ –, again, one would not thereby cause the generation or corruption of anything. There would never be anything *there* but the snake (the snake that is also a bracelet).

As for the House of the Wise Pig Argument (van Inwagen 1990, 130), insofar as the example of the ‘house of the Wise Pig’ can really be said to figure in an *argument*, the ‘conclusion’ of that argument is no more than this:

²⁴Note that this thesis contradicts constraint (6).

Suppose one's Answer to the Special Composition Question entails that there are houses²⁵ if things are arranged in certain ways – that is, that it entails that, for any xs , if those xs are arranged housewise at t , then they compose something, to wit a house, at t . And suppose one further believes that these 'houses' persist through time and can retain their identity through a change of parts – the removal or replacement of a brick, say, or the addition of a new wing. It is very difficult to formulate a metaphysic of houses with these properties that is consistent with both endurantism and the law of the excluded middle. (But it is not at all difficult to formulate a metaphysic of houses with those properties that is consistent with both *perdurantism* and the law of the excluded middle. This is one of the stronger arguments for perdurantism.) Endurantists can avoid having to confront these difficult problems if they adopt an Answer that does not permit the existence of houses.

And this conclusion by no means logically implies the thesis that there are no houses – or even that there are no non-living houses. Perhaps the problem of formulating an Answer that incorporates both endurantism and the law of the excluded middle *and* permits the existence of persisting non-living houses, although indeed difficult, can be solved. (And if a problem can be solved, that casts doubt on the thesis that a metaphysic that enables one to avoid that problem thereby enjoys an important advantage over any metaphysic that does not enable one to avoid that problem: the costs of the 'avoidance' metaphysic might be very high.²⁶) Or perhaps the law of the excluded middle requires amendment. Or perhaps perdurantism is the correct metaphysic of persistence.

12.4.4 *The Merits of an Answer That Assumes the Existence of Artifacts*

In Sect. 5.3, the authors give reasons for believing in the existence of artifacts. They do not, however, propose an Answer to the Special Composition Question that permits the existence of artifacts. But suppose they had. Let us call their imaginary Answer "A". Now compare this imagined answer A with my Proposed Answer:

($\exists y$ the xs compose y) if and only if the activity of the xs constitutes a life (or there is only one of the xs). (van Inwagen 1990, 82)

The most the authors' arguments could establish is this:

²⁵ By "houses" in the present discussion I mean houses that are normal artifacts. I have read at least one science-fiction story in which people inhabit houses that are living organisms. (Of course "houses" here stands in for any sort of normal artifact.)

²⁶ It would be nice to be able to avoid the problem of distinguishing between a thing's essential and its accidental properties. But that is not a persuasive argument for adopting *Spinozism*, the doctrine that everything has *all* its properties essentially.

Answer A permits the existence of (normal) artifacts and the Proposed Answer does not. Permitting the existence of artifacts is a ‘good-making’ property of an Answer. Therefore, Answer A has a good-making property that the Proposed Answer lacks.

Suppose they do establish that thesis. Still, the Proposed Answer, might, all things considered, be superior to Answer A. It is of course impossible to know whether this is the case – for the simple reason that Answer A is a creature of fiction. We know no more about it than that it permits the existence of artifacts. An Answer might permit the existence of artifacts but, say, require a counterpart-theoretical account of modal discourse *de re* (e.g., “Descartes could have lived longer than he actually did”). That is to say, in addition to this one good-making property, Answer A might have all manner of ‘bad-making’ properties – and the Proposed Answer might have only one bad-making property (i.e., that of not permitting the existence of artifacts). We really do not know, because we do not know what Answer A *is*.

Perhaps the authors will reply that not permitting the existence of artifacts is a *decisive* bad-making property of proposed Answers to the Special Composition Question – that is, that a proposed Answer that has that property, however many good-making properties it may have, *must* be regarded as unsatisfactory. I can say only that in my view their arguments fail to establish that very strong conclusion. I anticipated the arguments of this section in a paragraph that spans pages 127 and 128 of *Material Beings*. But perhaps I ought to have devoted more than one paragraph to that topic.

12.4.5 *Collective Functions*

I will close by replying to just one of the arguments of Sect. 5.3 of the authors’ paper; it may serve as a model for how I should reply to each of the others.

Our worry is that if creation consisted simply in a rearrangement of matter, it would be far from clear how to retain explanations that rely on the notion of function: If van Inwagen is correct, material artifacts are nothing but virtual objects – that is, objects that only seemingly exist. It is simply not clear how virtual objects could qualify as function bearers in the required sense: How can a virtual object have the function of boiling water? (Sect. 5.3.3)

My wife once had the interesting experience of watching Vladimir Putin walk from a door in one of the interior walls of the Kremlin to a door in the opposite wall. The President was surrounded at each moment of this short journey by three large, muscular, alert-looking men. Had those three men a function? Pretty evidently they did, and pretty evidently it was to protect their charge from assault. It is therefore possible for a plurality of things to have a function – or if that sounds as if I am treating “plurality” as a general term that refers to objects of a certain kind, I will put my point in these words: “It is possible for there to be *xs* that are two or more in number and which *collectively* have a function.” If we say, “The function of the three bodyguards was to protect Putin from assault”, we do not need to assume (or, for that

matter, to deny) that those three men had a mereological sum – an individual composite thing whose function was to protect Putin from assault.

“How can a virtual object have the function of boiling water?” – or let us rather ask, “How can a virtual object have the function of being a container in which water can be boiled?” Well, it cannot, not ‘in the strict and philosophical sense’, for the plain reason that, ‘in the strict and philosophical sense’, there *are* no virtual objects. (I accept the authors’ statement that something that is not *there* cannot have a function.) But a multitude of things arranged container-wise can *jointly* or *collectively* have the function of confining some water to a certain region of space while it is being heated till it boils. (Cf. the paragraph that starts at the bottom of page 117 of *Material Beings*.²⁷).

12.5 Fictional Entities: Reply to E.-M. Jung and F. Pellet

12.5.1 *Different Kinds of Theoretical Entities*

I should say, first, that in my use, “theoretical entity of physics” and “theoretical entity of literary criticism” are epistemological, not ontological, categories. The theoretical entities of a science or study or discipline are entities we *know about* in a certain special way. The theoretical entities of, for example, physics belong to a wide variety of ontological categories. Electrons and protons and neutrons are certainly theoretical entities of physics, and, whatever ontological categories they may belong to, they are certainly *causal* things, things with causal powers. (I once heard the Nobel laureate Sheldon Glashow exclaim, “A proton is a *thing* – like a *rock!*”) But other theoretical entities of physics are abstract objects, universals. Masses, for example – items such as the rest mass of the electron, the mass of the prototype kilogram in Sèvres, and the mass of Jupiter. Take the latter two masses: the mass of Jupiter is – is *numerically identical with* – the mass of the prototype kilogram multiplied by the number 1.898×10^{27} . The mass of Jupiter is (no doubt) also the mass of other objects to be found at various places in the cosmos. This mass is not a physical thing with causal powers²⁸ (like an electron or a proton or Jupiter itself). It is not some sort of *component* of Jupiter, some object such that it is likely that there is another gas giant somewhere in the vastness of the Hubble universe that has a component that is of the same ontological kind, numerically distinct from Jupiter’s mass but ‘equal’ to it; no, the mass of some other planet could well be numerically identical with the mass of Jupiter, could well be one and the same abstract object (to

²⁷In which, I am sorry to say, the word ‘sidereally’ is misspelled – or would be if there were such a word.

²⁸The mass of Jupiter *is* a causal power, but it does not *have* causal powers. Compare: 10 cubic meters *is* a volume; it does not *have* a volume. Only spatially extended objects (and perhaps regions of space) *have* volumes, and 10 cubic meters is not a spatially extended object or a region of space.

wit, 1.898×10^{27} kg) – just as the polar diameter of some other planet could well be numerically identical with the polar diameter of Jupiter, could well be one and the same abstract object (to wit, 1.33709×10^8 m). Note also that it is entirely possible that there are masses that nothing has; it could be that there is nothing in the universe that has exactly twice the mass of Jupiter. In that case, the mass 3.796×10^{27} kg *exists* and is a perfectly good example of a mass, but it is a mass that is not the mass *of* anything. It is evident, therefore, that masses are abstract objects of some sort; and it is equally evident that electrons are not abstract objects.

By a theoretical entity of some discipline, I mean only an entity (i) whose existence is ‘displayed’ by quantificational analysis of the theoretical discourse of that discipline – and *only* of that discipline, and (ii) is such that the fact that its existence is so displayed is the *only* reason there is to believe that it exists. (The qualification “and *only* of that discipline” is meant to block such counterintuitive consequences as that mathematical entities are theoretical entities of physics. But more needs to be said about this – something must be said to take account of the obvious fact that some disciplines involve applications of other disciplines: physics and chemistry and astronomy involve applications of mathematics; chemistry and astronomy involve applications of physics, and so on. I will not attempt to resolve this difficulty within the scope of an incidental remark.) Mrs. Gamp is indeed, ontologically speaking, nothing at all like an electron – but neither is the rest mass of the electron, ontologically speaking, anything at all like an electron.

12.5.2 *Having vs Holding Properties*

I found note 5 of the authors’ paper so deeply puzzling that I feel compelled to discuss it at greater length than might seem appropriate for a note. Here is the text of the note:

Van Inwagen (1977, 305) says here, for example, that fatness is ascribed to Mrs. Gamp. But, this is a misleading way to put things, because fatness is ascribed to a *character in a novel* to which is also ascribed the property of being called “Mrs. Gamp”. Saying that fatness is ascribed to Mrs. Gamp is precisely committing the mistake of thinking that Mrs. Gamp exists as someone who has properties like “being a character in a novel”.²⁹ If we follow van Inwagen’s line of thought here, then we should say, strictly speaking, that there *is* something, an entity, which *has* the property of being a character in a novel and which *holds* the properties of being called “Mrs. Gamp” and being fat. But directly saying that Mrs. Gamp *has* the property of being a character in a novel is a mistake. To summarize, there is the unfortunate tendency to say that Mrs. Gamp is a fictional character, while one should say that there is a fictional character called “Mrs. Gamp”. Note that van Inwagen (2003, 146) is more cautious by replacing “Mrs. Gamp” by “fictional characters”.

²⁹What is this mistake? Unless it was the authors’ intention to come down very hard on the words “someone who”, I do not know. I certainly affirm that Mrs. Gamp is *something that* has properties like “being a character in a novel.”

I can make very little of this. I will say some things that may be relevant. First, a fictional character's "holding" a property is the same as that property's being ascribed to that character – but with this qualification: the relation expressed by "the character x holds the property y " is a binary relation, and the relation expressed by "the property y is ascribed to the character x at the place z " is, of course, a ternary relation. The third term in the latter relation, the "place", is a story, a chapter, a passage, or something of that sort. This third term allows us to say things like "Youth is ascribed to the hero in the first chapter of the book and old age is ascribed to him in the final chapter" – or, as one would say in an ordinary literary conversation: "The hero is young in the first chapter of the book, and old in its final chapter."

But since we are not particularly interested in the expressive advantages of ascription over holding in the present context, let us use the simpler language of holding:

<i>Mrs. Gamp has</i>	<i>Mrs. Gamp holds</i>
being a character in a novel	fatness
being called "Mrs. Gamp"	being named "Mrs. Gamp"
holding fatness	being called "Mrs. Gamp"
	being a person made of flesh and bones

Mrs. Gamp *has* the properties I have listed in the same sense as that in which Vladimir Putin *has* the properties "being slight of build", "being named 'Vladimir Vladimirovich Putin'", and "being a person made of flesh and bones." Mrs. Gamp *has* these properties in the same sense as that in which iron *has* the property of being ferromagnetic, the number six *has* the property of being a perfect number, and the Taj Mahal *has* the property of being white.

Now what about "being named 'Mrs. Gamp'" and "being called 'Mrs. Gamp'"? What is the distinction I mean to be making? Is not "Mrs. Gamp" a name? – a *proper* name? Well, yes – in the sense in which philosophers of language use "name" and "proper name". In *that* sense, "J.F.K." is a (proper) name of the 35th president of the US, and " π " is a (proper) name of the ratio of the circumference of a circle to its diameter. But no one would say ("except as a joke", as they used to say in Oxford) that the 35th president was *named* "J.F.K." or that the ratio of the circumference of a circle to its diameter was *named* " π ". The statesman and the irrational number are (respectively) *called* "J.F.K." and " π ", yes. But those are not their *names* (not, at any rate, in the sense in which "Vladimir Vladimirovich Putin" is the name of the president of Russia).

Now *in the story* Mrs. Gamp has a name in the same sense of "name" as the sense in which Putin has a name in reality – it is "Sarah Gamp" or "Mrs Gamp" or "Sairey Gamp" (as you will). But then *in the story* Mrs. Gamp is a *real, existent, flesh-and-blood, non-fictional human being*. That is to say, just as she *holds* the property "being a real, existent, flesh-and-blood, non-fictional human being," she holds the property "having the name 'Sarah Gamp'". But *we're* not "in the story". Why do we call her "Sarah Gamp"? Because, as I have said:

I would suggest that the *only* reason “Mrs. Gamp” denotes a certain creature of fiction is that that creature of fiction satisfies the open sentence “A (being named ‘Mrs. Gamp,’ *x*, *Martin Chuzzlewit*).” That is, we have embodied in our rules for talking about fiction a convention that says that a creature of fiction *may* be referred to by what is (loosely speaking) the name it has in the story. It is owing to a similar convention that we use *personal* pronouns in connection with entities that are literally *not* persons: we call Mrs. Gamp “she” because, though she is not a woman, the property of being a woman is ascribed to her.³⁰

We do not call a certain fictional character “Sarah Gamp” because Dickens, sitting at his writing-desk at some moment in the 1840s said, “I will give the character I am about to introduce the name ‘Sarah Gamp’”. (He *may* have said something like that to himself – and he may not have. We have no idea, really.) That is, we do not call her “Sarah Gamp” because Dickens named her “Sarah Gamp” in the way he (or he and his wife) named one of his sons “Frank”. Nor do we call her that because some reader of *Martin Chuzzlewit* said something along the lines of, “I propose that we all refer to the main comic villainess of this novel as ‘Sarah Gamp’ at an early meeting of the Charles Dickens Fan Club. We call her “Sarah Gamp” – we *call* her “Sarah Gamp” – because she *holds* the property of being *named* “Sarah Gamp”. We call her “the main comic villainess of *Martin Chuzzlewit*” because she *has* the property of being the main comic villainess of *Martin Chuzzlewit*.

12.5.3 Creationism Concerning Theoretical Entities

The authors are right to suppose that I reject creationism. (At any rate, I *now* do. In my earlier writings on the ontology of fiction, I made some rather half-hearted attempts to remain neutral as between creationism and anti-creationism.) The difficulties they raise for anti-creationism are largely the standard ones. They apply equally to anti-creationism about novels – for I certainly think that, say, *War and Peace* exists in all possible worlds. That, however, does not mean that that great novel has no ‘historical presence’ in our actual world or no connection with Tolstoy’s literary labors. After all, Tolstoy *wrote* it. (If Tolstoy had died in infancy, *War and Peace* would still have existed; but it would have been unwritten. Would it have been a *novel* – an unwritten novel? A matter of definition, I suppose – but I’d prefer to say that it became a novel only when Tolstoy had written it.) He *published* it. He *caused* it to be accessible to readers.

The situation is somewhat similar with possible chess games. According to one estimate (probably as good as any) there are about 10^{123} of these. Obviously, most possible chess games will never be *played*. Obviously, they are abstract objects of some sort. But that does not mean that history is irrelevant to chess: *one* of those unimaginably many possible games was in fact *actually played* by Bobby Fischer and Boris Spassky on 11 July 1972 in Reykjavik. And those two gentlemen did not

³⁰Van Inwagen 1977, 307. The quotation is partly from the text and (starting with “That is [...]”) partly from note 8. See also note 7.

reach into some Platonic ‘chess heaven’ and pull an abstract object down into the world of flux and impermanence. The fact that this abstract object, this game, was actually *played* was the outcome of a very complex historical causal process.

Or consider the proposition that the number of cut diamonds with a weight of more than 1000 carats is smaller than the number of women who are more than two meters tall. Like every proposition, it is an abstract object and exists in every possible world – but the fact that it got mentioned in this Reply is the result of a complex causal-historical process. (One that went on primarily in my own brain: I set myself the task of imagining a bizarre but true proposition that – to a near certainty – no one else had ever thought of.) The authors ask:

How could an abstract object, for instance, be regarded as a stereotype of the poorly competent nurse in the early Victorian era if it is of an eternal nature? And how could we think of Mrs. Gamp as being independent of human mind? In van Inwagen’s picture, Mrs. Gamp has always been there. She has never been created by Dickens. We could at least say that he *detected* or *discovered* her. (Sect. 6.5.1, emphasis in the original)

I would ask: if something is a stereotype (of any sort of thing whatever), what could it be *but* an abstract object? Can one touch a stereotype? Can one point to its location in space? Difficult philosophical problems do indeed present themselves when philosophers consider the question of the relation between abstract objects and the human *Lebenswelt*. Mathematics abounds with examples. Brouwer’s Fixed-point Theorem is an abstract object if anything is. How then can it first have been proved, independently by L. E. J. Brouwer and Jacques Hadamard, in 1910? And the proofs of Brouwer and Hadamard are themselves abstract objects. How then can they both depend historically on a long series of mathematical discoveries in topology and the theory of differential equations in the nineteenth and early-twentieth centuries? How can we think of these proofs as being independent of the human mind? We would certainly *not* say that those two great mathematicians “detected” their proofs (they did not peer into a mathematiscope and observe them floating in mathematical space, like an astronomer searching a section of the sky and coming upon a hitherto unknown comet). We *might* say that they “created” their proofs, but such language is in fact rarely if ever used. To say that a mathematician has “discovered” a proof is a fairly common way of speaking – but discovering a proof is nothing at all like discovering a comet or discovering a First-Folio Shakespeare in the library of an obscure country house.

In short, the problems with which the authors’ confront the anti-creationist in the ontology of fiction have fairly precise analogies in most or all cases in which some field of theoretical enquiry involves reference to or quantification over abstract objects.

12.5.4 Two Further Points

I do not see how anyone could regard Yagisawa's thesis (discussed in Sect. 6.5.2) as even faintly plausible. Here are two theses:

The half-life of a free neutron is about 10.8 min.

There is at least one mystery novel whose text is in the form of a first-person-singular narrative and in which it is revealed at the end of the novel that the murderer is the narrator himself.

Each sentence expresses a propositional truth. Each is a statement of an objective fact. The second is a statement that belongs to what I am calling "literary criticism". Whatever someone's purpose in writing this sentence might be, it would certainly *not* be to enhance any reader's aesthetic experience. The goal of literary criticism, (supposing that one can state any non-trivial general thesis about literary criticism without writing a monograph) is to *understand* various aspects of literature – to understand how authors produce their effects, to understand how literature and culture interact, to achieve an *intellectual* understanding of these and many, many other matters.

[T]he criterion for the justification of assuming that fictional objects are existent is obviously a *contingent and historically dependent criterion*. We are justified in assuming that exactly these objects, which belong to the theoretical discipline of literary criticism and its traditions, do exist. What would happen, for example, if the golden mountain became an important part of a novel and, thus, a matter of debate in literary criticism? And what would we say about the existence or non-existence of these fictional characters and objects in novels, which for contingent reasons have never been discussed by literary critics? (Sect. 6.5.2, emphasis in the original)

The purpose of Quine's "criterion of ontological commitment" is not to enable one to discover what exists *simpliciter*, but to enable one to discover what one's statements and theories *imply* the existence of. If I am right, then this criterion entails that people who say things like "The characters in some nineteenth-century novels are presented with a greater wealth of physical detail than is any character in any eighteenth-century novel" are committed to the existence of fictional characters. But fictional characters are among the primary, and the most plausible, examples of "non-existent things" or "things that have no sort of being whatever" that Meinongians and neo-Meinongians are able to offer. Do not suppose that Meinong arbitrarily made up the example of the golden mountain. That prominence was, rather, a fictional mountain well known to his contemporary readers, an inhabitant of *Nichtsein* (so Meinong supposed) that Meinong could rightly assume had been the object of the individual thoughts of a great number of people. (The fiction in which it figures is not a novel or short story but a *Märchen* – Grimm 92, "Der König vom goldenen Berg".) My ontology of fiction is designed to enable Quineans to say, "Yes, Meinong, the words 'the golden mountain' do have a referent, but it is not a *non-existent mountain*, as you have supposed – it is not a non-existent thing that *has*

the property ‘being a (golden) mountain’; it is rather an *existent* thing that *holds* the property ‘being a (golden) mountain’.”

12.6 Fictional Entities: Reply to K. Gregor, J. G. Michel and S. Neuß

Much of what I said in response to the essay of Jung and Pellet is relevant here, but I will try not to repeat myself.

12.6.1 *Creation of Abstract Entities*

In early work on the ontology of fiction (van Inwagen 1977, 304), I tried to maintain a position of strict neutrality as between theories of fictional characters according to which fictional characters are abstract objects³¹ and theories according to which authors literally bring fictional characters into existence. (“But why not say both that fictional characters are abstract objects *and* that authors bring them into existence?” A good question – to which I shall return.)

This neutrality is no longer possible for me. For I now accept a general ontology that recognizes only two ontological categories:

- objects that fall under the following general terms: “relation”³², “abstract object”, “anetiological or non-causal object”, “universal”.
- objects that fall under the following general terms: “substance”, “concrete object”, “causal object”, “particular”, “individual”.

I now maintain, moreover, that all objects in the first category are necessarily existent. (It will be observed that there are some conspicuous absences from the two lists of general terms that I have used to define the two categories: “set”³³ and “event”, for example. Their absence is a topic for another place.) And if these are the ontological categories, and if fictional characters exist, they certainly have to be placed in the first category. It follows immediately that I am committed to the position that fictional characters are not created by their authors – and this for two reasons: they are necessarily existent; “creation” is a causal relation. My commitment

³¹ But in one place I slipped and described them explicitly as abstract objects.

³² “Relations” include propositions (0-term relations), properties or attributes (unary relations), proper relations (binary, ternary, ... relations), and multigrade relations (“teach at the same university”, “are members of the CDU”, ...).

³³ Naturally, it is incumbent upon me to explain how we can ‘do without’ sets. But, assuming that I can do that, I need not reply to the argument, “But if anything can be created, then sets are abstract objects that can be created – for to bring x and y into existence is to bring $\{x, y\}$ into existence.”

is due not to anything having to do with the ontology of fiction in particular, but by very general features of my philosophical system if I may be said to have anything so grand as a philosophical system.

In my reply to Jung and Pellet, I defended the position that, whatever counterintuitive consequences “anti-creationism”, may have, they are not confined to the ontology of fiction. The same “counterintuitive consequences”, *mutatis mutandis*, arise when one considers (for example) the ontology of mathematical proof – and the ontology of games of chess.

12.6.2 *Vulcan*

[...] Vulcan could be proved to exist, though obviously not as a spatial entity, but as a creature of myth, whereas, e.g., the moon’s existence would have to be denied, for the moon is neither a simple nor a living organism. But this conclusion contradicts our everyday beliefs.

I will remark first that when the authors say, “Vulcan could be proved to exist as a creature of myth,” I *think* they are speaking of the once-hypothesized inter-Mercurial planet, and not the Roman god after whom it was named. But I see no reason to think that there is a creature of myth that holds the property “being an inter-Mercurial planet with a mass greater than one sixth that of Mercury”. (I would ask: in the mythology of what age and culture does that planet figure? A once-popular scientific hypothesis that turns out to have been false is not a *myth*.) And if I am mistaken and in that sentence the authors are using “Vulcan” to refer to the Roman counterpart of Hephaestus, I do not see the function of the statement, “Vulcan could be proved to exist as a creature of myth” in their argument. As to the statement that the non-existence of the moon “contradicts our everyday beliefs” – well, I concede that many philosophers are convinced that it is true. But, of course, in *my* view it is false. See §10 of *Material Beings* (van Inwagen 1990) – whose title “Why the Proposed Answer to the Special Composition Question, Radical Though It Is, Does Not Contradict Our Ordinary Beliefs” well describes its content (see also van Inwagen 2014a).

12.6.3 “*Sherlock Holmes Does Not Exist*”

In my view, there is no one proposition that is expressed by the sentence “Sherlock Holmes does not exist” when it is spoken in what I like to call the ordinary business of life. It might express this proposition:

No one has all the properties held by Sherlock Holmes.

(Context: A frustrated police detective says, “Only Sherlock Holmes could solve this case, and, unfortunately, Sherlock Holmes does not exist.” The detective knows that the members of her audience are perfectly well aware that Holmes is a fictional character.)

Or it might express this proposition:

The words “Sherlock Holmes” did not enter into human discourse by being given to a child in baptism or by being inscribed on a birth certificate; their linguistic origin rather lies in the fact that the property “being named ‘Sherlock Holmes’” is held by a certain fictional character.³⁴

(Context: a naïve tourist in London is trying to find 221B Baker Street. He asks a police officer where exactly the building with that address lies – “You know, where Sherlock Holmes lived”. And he is told, “Sir, I am sorry to have to tell you that Sherlock Holmes does not exist and never did. He’s just a bloke in a story made up by someone called Conan Doyle.”)

Or, finally, it might express the proposition that there is no such fictional character as Sherlock Holmes.³⁵ (In several essays, I have imagined a circumstance in which someone might assert that proposition.³⁶)

12.6.4 *Some Answers to Questions in the Dialogue*

I record my own replies to several of the questions posed to my fictional counterpart.

First question

Do you think Sherlock Holmes is still somewhere out there, waiting for new riddles to solve? (Sect. 7)

No, obviously not. The concept of location is not applicable to Holmes. One might as well ask, “Do you think the conjecture Christian Goldbach put forward in a letter to Euler on 7 June 1742 is still somewhere out there, waiting for someone to come along and prove it?” Of course, one might try to interpret the authors’ question in terms of the properties Holmes “holds” rather than the properties he has (an interpretation encouraged by the words “waiting for new riddles to solve”). We can imagine one present-day Holmes aficionado asking another, “Where was Holmes on the day of Watson’s wedding?”, and, for all I know, the question has a determinate answer. But if the questioner went on to ask, “And where do you think Holmes is *today*? Is he still somewhere out there waiting for new riddles to solve?”, the

³⁴ See the discussion of fictional names in the Reply to Jung and Pellet’s essay.

³⁵ Or perhaps not precisely those three propositions but certainly propositions intimately related to them. Here’s an analogy—an analogy based on the following well-known anecdote. A haughty French royalist said to an English visitor, “Your House of Lords is unworthy of the name. No English nobleman (so-called) of the present day would be permitted to ride in the same coach as a king of France.” The Englishman replied, “There is no king of France.” Question: What is the relation between the proposition the Englishman asserted and the proposition that either no male now reigns over France or more than one male now reigns over France? It is plausible to suppose that it is this relation is not strict identity, but the two propositions are nevertheless certainly very intimately related.

³⁶ See, e.g., van Inwagen 2003, n. 14.

answer could only be something like, “Are you serious? If Holmes were still alive, he would be almost 170 years old. If he is anywhere, it is in a coffin. He’s certainly not in a position to solve any riddles.”

Second Question

What exactly does “holding” mean? Could you give a more precise definition of a term so central to your approach? (Sect. 7)

Well, I can certainly give what I regard as an excellent explanation of the meaning of the term, even if it is not a definition in the strict and Chisholmian sense.³⁷ Consider these two columns of properties:

having been born in 1850	having been born in 1066
being a consulting detective	being a butcher
being a heavy smoker	being an anti-tobacco fanatic
being an excellent violinist	suffering from amusia
living in London	living in Moscow

There is some salient and important relation that Holmes bears to all the properties in the first column and does not bear to any of the properties in the second column – and, in fact, he bears this relation to the negation or complement of all the properties in the second column. And, what is more, he bears this relation to none of the properties in either of *these* two columns:

being the son of a woman named “Alice”	not being the son of a woman named “Alice”
having a mole on his left shoulder	not having a mole on his left shoulder
having an Irish second cousin	not having an Irish second cousin

Whatever this salient and important relation is, it is not *having*, and this for at least two reasons: (i) Holmes is an abstract object and thus has none of the properties in the first column of the first list; (ii) for every property, everything has either that property or its complement; and thus, for each property in the first column of the second list, everything has either that property or the property opposite it in the second column³⁸ – and yet Holmes bears the “salient and important” relation in question to none of the properties in either column of the second list.³⁹

“Holding” is my name for the salient and important relation that Holmes bears to the properties in the first column of the first list and bears to none of the other properties in either column of either list.

³⁷The “Cartesian” analogy was a mere incidental remark.

³⁸Holmes, the Riemann curvature tensor, and I each *have* all the properties in the second column of the second list.

³⁹And here is a third reason: It is possible for a fictional character to hold inconsistent properties; and, of course, nothing can *have* inconsistent properties.

Third Question

Imagine that each single copy of the novels about Sherlock Holmes has vanished and nobody remembers the stories anymore. Would that mean that the fictional character of Sherlock Holmes, too, doesn't exist any longer? (Sect. 7)

Yes and no. The abstract object that Holmes is would still exist (in the following sense of “still exist”: if fictional characters are indeed abstract objects, the concept “ceasing to exist” has no application to them), but he, or rather *it*, would no longer be a fictional character. That is to say, if fictional characters are abstract objects, then “fictional character” is a *phase sortal* – like “philosopher” and “kitten” and “Chancellor of the Federal Republic of Germany”. That abstract object would still, of course, have the modal property “is possibly a fictional character” (the number of objects that have that modal property is vastly greater than the number of objects that have the property “is a fictional character”) and it would have the historical property “was once a fictional character”. The status “fictional character” may be instructively compared with the status “theorem”. The mathematician Paul Koebe is reported to have said, when viewing Leonardo's *The Last Supper*, “How sad! This painting will pass away, while my theorem on the uniformization of analytic functions will endure forever!” Well, the *mathematical proposition* that Koebe proved will “endure forever” – in that the concept “ceasing to exist” has no application to propositions. But it became a *theorem* only in 1907 (when Koebe proved it), and (at any rate, *I* find this a plausible thing to say) it will cease to be a theorem when it is no longer remembered by anyone. In possible worlds in which there are never any mathematicians, it exists but is never a theorem. (In fairness to Koebe, it may well be that he really meant only that a mathematical proposition's having the status “theorem” does not depend on the existence of any particular physical inscription of the theorem and its proof.)

Fourth Question

Can you provide us with some kind of a minimal criterion for the genesis of fictional creatures? (Sect. 7)

If I contend, as I do, that fictional characters are abstract objects (and that abstract objects are one and all necessarily existent and anetiological), then the question that should be addressed to me is, “Can you provide us with some kind of a minimal criterion for the acquisition of the status ‘creature of fiction’ by an abstract object that has the property ‘is possibly a creature of fiction’?” This is indeed a difficult question, but it is not the same question as the one the authors have asked. I will mention a text that the philosopher who addresses this question must regard as embodying the intuitive image of “authorial creation” that everyone has – some famous words spoken by Theseus in Act 5 of *A Midsummer Night's Dream*:

And as imagination bodies forth
The forms of things unknown, the poet's pen
Turns them to shapes and gives to airy nothing
A local habitation and a name.

Fifth Question

In [*Hamlet*], another play occurs, namely the play *The Murder of Gonzago*. In the play *Hamlet*, *The Murder of Gonzago* is a play Hamlet has performed in order to determine his uncle's innocence or guilt in the death of his father. Apart from that, it obviously centers on the murder of someone named "Gonzago". Well, from my perspective both Hamlet and Gonzago are creatures of fiction. Do you see a difference between them? Could we, for instance, say that Hamlet is a fictional character, while Gonzago is a fictional fictional character? (Sect. 7)

The example is Kripke's (2013, 59–62).⁴⁰ Actually, it is not a very good example, as you will see if you read *Hamlet*, Act III, Scene 2. Here is a much more straightforward one. Scheherazade is a fictional character, a character in *A Thousand and One Nights*. In that curious work, Scheherazade tells stories – the story of Aladdin and the lamp, the story of Sinbad the Sailor, and many more. Now if Scheherazade is a fictional character, what is Aladdin? The answer to this rhetorical question is: He is a creature of fiction, but not a fictional character – not at least if a fictional character is a creature of fiction that holds the property "being a human being". (So, for example, the palace in which Scheherazade tells her tales is not a fictional character, although it is a fictional palace – that is, a creature of fiction that holds the property "being a palace".) Aladdin is a *fictional* fictional character – just as the authors have suggested (and this is also Kripke's position). That is to say, Aladdin does not *have* but *holds* the property "being a fictional character" – a property that Scheherazade *has*. Scheherazade, moreover, holds the property "being a human being" (she is a creature of fiction who holds the property "being a human being", and is thus a fictional character), while Aladdin holds the property of holding the property "being a human being" (he is a creature of fiction who holds the property of holding the property "being a human being", and is thus a fictional fictional character.) Scheherazade is a fictional woman; the palace in which she tells her tales is a fictional palace; Aladdin is a fictional fictional character: there is no such fictional character as Aladdin, but *in the story*, in *A Thousand and One Nights*, there *is* a fictional character who holds such properties as "being named 'Aladdin'" and "commanding the services of a powerful djinn" and so on. If you find this iteration of "fictional" confusing, I am sorry to have to tell you that the ontological realm in which Aladdin lives and moves and has his being occupies only the first level of iteration, for in *A Thousand and One Nights* there are stories within stories within stories ... And thus there exist (without qualification) fictional fictional fictional characters, fictional fictional fictional fictional characters ... All are creatures of fiction, but (this is a tautology worth attending to) only fictional characters are fictional characters – that is to say, a creature of fiction is a fictional character only if it holds the property "being human."⁴¹ Creatures of fiction that hold the property of holding the property "being human" are not fictional characters – and *a fortiori* creatures of

⁴⁰ Kripke's ontology of fiction and mine are identical – although their genesis is independent; cf. van Inwagen 2014b.

⁴¹ Or at any rate, the property "being a rational being". Tolkien's elves and Dr Frankenstein's creature and E.T. are fictional characters, I suppose.

fiction that hold the property of holding the property of holding the property “being human” are not fictional characters.

Sixth Question

[I]f you have the opportunity to visit the Reichenbach Falls you will take note of an interesting plaque on the ledge [...], which says: “At this fearful place, Sherlock Holmes vanquished Professor Moriarty, on 4 May 1891”. Well, obviously, this is an assertion. Now, what do you think: has Sherlock Holmes been there? (Sect. 7)

Consider a hobby horse. Its purpose is to be a prop in a game of pretense played by a child. The child who so uses it is ‘pretending’ to ride a horse – but not in the sense in which a confidence trickster who is attempting to sell her victim the Brooklyn Bridge is pretending to be the current legal owner of the Brooklyn Bridge. For no one is deceived, neither the child herself nor the adults who witness her frolicking about with the hobby horse between her legs: She is not attempting to produce in anyone – herself included – the belief that she is *actually riding a horse*.

The plaque in Meiringen is also a prop in a game of pretense: Those who placed the plaque beside the Reichenbach Falls and the tourists who since its emplacement have paused to read its inscription (at least those among the tourists who are aware of the content of the fictions “The Final Problem” and “The Adventure of the Empty House”) are pretending that the Holmes fictions are real historical memoirs written by a British physician named John Watson (they are pretending this in the same sense that the child astride the hobby horse is pretending to ride a horse) and the plaque is (and was designed and constructed to be) a prop in a game of pretense played by visitors to the Reichenbach Falls – in precisely the same sense as the hobby horse is (and was designed and constructed to be) a prop in a child’s game of pretense.

Now consider a more elaborate and up-to-date child’s toy: a flat surface covered with iconic representations of buttons and dials and switches that is ‘supposed to be’ the control board of a space ship – in the sense of “supposed to be” in which the hobby horse is supposed to be a horse. One prominent feature of this toy is a large red dot that is ‘supposed to be’ a button. Above this ‘button’ there is inscribed,

WARNING: SHIP SELF-DESTRUCT BUTTON

and beneath it there is inscribed

IF THIS BUTTON IS PRESSED, ANTI-MATTER CONTAINMENT WILL BE BREACHED AND THE SHIP WILL BE DESTROYED.

“Well, obviously, this is an assertion.” No, it is not. It is an inscription that pretends to be an assertion – just as the red dot, which is not a button, pretends to be a button.⁴² The same is true of the words inscribed on the plaque in Meiringen.

⁴²That is to say: the dot was placed on the toy control board by its manufacturers with the intention that children playing with the toy can pretend that the dot is a button.

As to the question whether Holmes was ever in Meiringen, the answer is obvious: He *has* no spatial properties whatever; he *holds* the property “having once been in Meiringen”.

12.6.5 *Implicit Characters*

A word about “implicit characters”. Platonism (sc. concerning fictional characters) *per se* entails neither that there are implicit characters nor that there are not. But some *versions* of Platonism have this implication. For example, Nicholas Wolterstorff’s version of Platonism⁴³ entails an *extreme* version of the proposition that all works of fiction contain implicit or ‘offstage’ characters, at least given some plausible premises concerning which propositions are “true in” a given fiction.⁴⁴ Suppose it is true in a certain fiction that every human being has a mother. Let this fiction be *The Sign of the Four*. It follows that it is true in that novel that Holmes has a mother. (Or consider the undeniable fact that the proposition “Holmes has a brother” is true in *The Sign of the Four*. Since it would be self-evidently true to any reader of *The Sign of the Four* that if two men are brothers, they have the same mother, “Holmes has a mother” is presumably true in that novel.) Wolterstorff’s ontology of fiction implies that if “Holmes has a mother” is indeed true in *The Sign of the Four*, then Holmes’s mother is a character in that novel – it implies, that is, that one of the characters in *The Sign of the Four* holds the property of being the mother of someone who has the property *F* – where *F* is the conjunction of all the properties that Holmes holds. But no such character ever, so to speak, *appears* in *The Sign of the Four*; nor does any ‘on stage’ character in the novel ever mention or refer to any of Holmes’s relations but his brother Mycroft. I regard this as an odd and counterintuitive feature of Wolterstorff’s ontology of fiction – a *prima facie* reason for someone who accepts Platonism to accept some version of Platonism other than Wolterstorff’s. (Prima facie: it might turn out that every other version of Platonism has consequences that are even more ‘odd and counterintuitive’ than the proliferation of implicit characters entailed by Wolterstorff’s version.)

⁴³For references and a summary of Wolterstorff’s theory, see van Inwagen 2003, 150–153.

⁴⁴For an analysis – a characteristically brilliant analysis – of the concept “truth in a fiction”, see Lewis 1978, 37–46.

12.7 Free Will: Reply to A. Savarino and A. Sock

12.7.1 Reformulating the Third Strand of the Mind Argument

The usual technical vocabulary of discussions of ‘the problem of free will’ is firmly entrenched in the authors’ paper. For the reasons I set out in “The Problem of Free Will Revisited”, I have grave reservations about that vocabulary. Accordingly, in what follows, I will set out a version of the *Mind* Argument (more exactly, of the “third strand” of the *Mind* Argument) that is stated in a vocabulary that I do find acceptable and examine the question of the applicability of the authors’ arguments to that version of the argument.

Let us say that an agent is *doubly able* at a moment t just in the case that:

That agent is at t deliberating between two mutually incompatible courses of action (the agent is trying to decide whether to lie or to tell the truth, for example) and is able to pursue each of them (the agent is, for example, both able to lie and to tell the truth – although not, of course, able both to lie and to tell the truth).

Let us say that *the outcome of an agent’s deliberations is undetermined* at the moment t just in the case that:

The agent is at t deliberating between two mutually incompatible courses of action X and Y , and

- in some possible world in which (i) the agent and the agent’s causally relevant environment⁴⁵ are in precisely the same states at t as those they are in in the actual world at t , and (ii) the laws of nature are the same as those of the actual world, the agent will decide to do X
- in some possible world in which (i) the agent and the agent’s causally relevant environment are in precisely the same states at t as those they are in in the actual world at t , and (ii) the laws of nature are the same as those of the actual world, the agent will decide to do Y .

And let us say that *the outcome of an agent’s deliberations is uniformly undetermined* at the moment t just in the case that:

- (i) The outcome of that agent’s deliberations is undetermined at t , and
- (ii) the outcome of that agent’s deliberations will be undetermined at every moment between t and the moment at which the agent’s deliberations ‘terminate’ in a (final, irrevocable) decision.

⁴⁵The agent’s causally relevant environment comprises those aspects of the agent’s environment that might have some sort of ‘causal input’ into the decision the agent reaches. If, for example, the agent is being interrogated by the police, and one of the interrogators says, “If you lie, you will be severely punished”, that would certainly be a part of the agent’s causally relevant environment. An argument between two people in an aircraft flying overhead at that moment would not be a part of the agent’s causally relevant environment – or not unless the agent could somehow hear it (by radio, perhaps).

(Imagine an agent A who now, an hour before noon, has begun to try to decide whether to do *X* or to do *Y* at noon. Suppose that the outcome of A's deliberations is now undetermined, but that at 11:30 it will either become determined that at some moment thereafter A will decide to do *X*⁴⁶ or become determined that at some moment thereafter A will decide to do *Y* – although of course it is now undetermined *which* of those decisions will become determined at 11:30. Then the outcome of A's deliberations is now undetermined, but it is not now *uniformly* undetermined – for there *will* be an interval of time during which A is still trying to decide whether to do *X* or to do *Y* at noon, but during that interval it will either be determined that A will eventually decide to do *X* or determined that A will eventually decide to do *Y*.)

Having these definitions at our disposal, let us proceed to address the following question:

Is it possible that, for some agent and some moment *t*, (a) the outcome of that agent's deliberations is uniformly undetermined at *t*, and (b) that agent is doubly able at *t*?

Let us consider a particular case. Suppose that one hour from now Alice will be asked an important question, a question that must be answered Yes or No, and that she has begun to try to decide which answer to give. And suppose that it is now uniformly undetermined whether she will decide to answer Yes or decide to answer No.

Suppose further that Alice *knows* that it is now uniformly undetermined whether she will decide to answer Yes or decide to answer No. Imagine that a friend begs her to promise that she will answer Yes (and that at the moment the friend makes this request it is uniformly undetermined how Alice will respond to it and that Alice knows this). Should Alice regard herself in a position to make such a promise?

It would seem that she should not. Consider first – as a sort of 'intuition pump' – a case that involves 'external' indetermination: Alice is about to throw a dart at a dartboard, and she knows that at the moment the dart leaves her hand it will be undetermined whether it will hit the left half or hit the right half of the dartboard. In that case, Alice is certainly not in a position to promise that the dart will hit the right-hand side of dartboard. If the outcome of a process or sequence of events (such as a cast dart's moving through space toward a dartboard) is undetermined, no one is in a position to make any promises about what its outcome will be.

Now suppose that an outcome-uniformly-indeterminate process takes place *within* Alice. Suppose, in fact, that that process is the very process of deliberation that will eventuate either in her deciding to answer Yes or in her deciding to answer No. And suppose that Alice knows that the outcome of her deliberations is uniformly undetermined. Would she not then be reasoning correctly if she said to herself, "It is uniformly undetermined whether I shall decide to answer Yes or decide to answer No – and, because it is uniformly undetermined what my decision will be, it will

⁴⁶"It is determined at *t* that at some moment thereafter A will decide to do *X*" is defined as "In every possible world in which (i) the agent and the agent's causally relevant environment are in precisely the same states at *t* as those they are in in the actual world at *t*, and (ii) the laws of nature are the same as those of the actual world, the agent will at some moment after *t* decide to do *X*."

'still' be undetermined how I shall decide to answer if I promise my friend to answer Yes.⁴⁷ If I promise to answer Yes, that promise *might* be followed by my deciding to answer Yes, but it *might* equally well be followed by my deciding to answer No. There is nothing more that can be said. If I make the promise, I might keep it and I might not keep it. Therefore, if I were to make the promise, then, for all I can say now, I should not keep it. Therefore, I am *not in a position* to make that promise."⁴⁸

Now if a person believes that he or she would be able to keep a certain promise if it were made⁴⁹, then that person is in a position to make that promise. And that is true even if the belief is false. If I believe that I should be able to give a lecture at the University of São Paulo next week if I promised to, then I am in a position to promise to give a lecture at the University of São Paulo next week even if (unbeknown to me) the stringency of Brazil's visa requirements renders me unable – promise or no promise – to give a lecture at the University of São Paulo sooner than three weeks from the present date. And it is also true if the promise in question is one the person would not dream of making. Perhaps I would not dream of promising to give a lecture at a place many thousands of kilometers distant from my home a mere seven days from now – nevertheless, if I believe that I should be *able* to do so if I promised to, I am in a *position* to promise to do so. (Not every promise one is in a position to make is a promise one would seriously consider making.)

Now let us return to Alice. If what I have said about the relations between ability, outcome-uniformly-indeterminate deliberation and being in a position to keep a promise is correct, then Alice may reason as follows:

1. It is uniformly undetermined whether I shall decide to answer Yes.⁵⁰
2. If it is uniformly undetermined whether I shall decide to answer Yes, then I am not in a position to promise to answer Yes.

⁴⁷This is precisely the reason for the distinction between the outcome of an agent's deliberations being "undetermined" and being "uniformly undetermined". For suppose that it is now undetermined which answer Alice will give, but it is also the case that *if* she promised to answer Yes, that promise would have the effect of rendering it determinate that she would answer Yes. (In that case, if she is in fact going to promise to answer Yes, whether she will answer Yes is not *uniformly* undetermined: it is undetermined *now* but it will become determined when she makes the promise.) If Alice knew that promising to answer Yes would render it determined that she would answer Yes, she *would* be in a position to promise to answer Yes, despite the fact that it is undetermined (*merely* undetermined) whether she will answer Yes. I am indebted to Michael Bratman for this important point.

⁴⁸Let us understand what a person's "not being in a position" to make a certain promise means – or, rather, what it does *not* mean. It does not mean that that person is *unable* to make the promise. We may imagine a confidence trickster who says to you: "I will send you a check for €10,000 tomorrow." In speaking these words, she promised to send you a check for €10,000 tomorrow – and she was therefore *able* to promise to send you a check for €10,000 tomorrow. But, we may further imagine, she was *not in a position* to make that promise, owing to the fact that she did not have even a tenth that sum at her disposal and knew it.

⁴⁹Perhaps I should add here the clause "and if that is a reasonable belief for that person to have". If that qualification is indeed necessary, it will make no difference to my argument.

⁵⁰In our 'official' vocabulary: "I am deliberating about whether to answer Yes or to answer No, and the outcome of my deliberations is uniformly undetermined."

3. If I should be able to decide to answer Yes if I promised to answer Yes, then I am in a position to promise to answer Yes.

Hence,

4. It is not the case that I should be able to decide to answer Yes if I promised to answer Yes.
5. If it not the case that I should be able to decide to answer Yes if I promised to answer Yes, I am unable to answer Yes.

Hence,

6. I am unable to answer Yes.

And, of course, by parallel reasoning, Alice may deduce that she is unable to answer No – for if it is uniformly undetermined whether she will decide to answer Yes, it must also be uniformly undetermined whether she will decide to answer No. Alice may therefore validly conclude that that she is not (at present) doubly able: she is unable to answer Yes, *and* she is unable to answer No – and this despite the fact that she is going to do one or the other. And she may further conclude that at *every* moment till she actually makes her decision she will be unable to answer Yes and unable to answer No.

“But suppose she answers Yes. Then she must have been *able* to answer Yes, since she *did* proceed to give that answer. And suppose she answers No. Then she must have been able to answer No, since she *did* proceed to give that answer. She was going to give one answer or the other – so she must have been able to give at least one of the two answers: the one that, in the event, she *did* give.”

There may be a sense of “able” in which it follows logically from “A *did* X” that “A was able to do X”. And we often do make unreflective use of this inference-form without specifying any particular sense of “able”.⁵¹ But if there is such a sense, it is not the sense that figures in questions of excuse from blame or fault – it is not the sense that the German word “*Können*” has in “*Sollen impliziert Können*” (“‘Ought to’ implies ‘is able to’”). If I randomly twirl the dials on a locked safe, and by sheerest chance, ‘hit’ the combination and the safe opens, it does not follow that I was able to open the safe: I *did* open the safe, but I was not *able* to – not, at least, in the sense of “able” that figures in “questions of excuse from blame or fault”. If a robber says, “Open the safe or I’ll beat you severely”, and I say “I’m not able to open the safe – I don’t know the combination”, what I have told the robber is true. And if, desperate to avoid a beating, I spin the dial at random several times and am extraordinarily lucky and the safe does open, that does not change the fact that I *did not have the ability* to open the safe.⁵²

⁵¹ I made use of it myself in this very Reply – in the “confidence trickster” example in note 48 (Sect. 12.7.1).

⁵² Suppose that at noon Alice decided to answer Yes. We should note that even if my reply to the Interlocutor contains a mistake, and—because she *did* answer Yes—Alice was able to answer Yes, it does not follow that she was able to answer No; that is, it does not follow that she was doubly able.

This argument for the conclusion that the fictional Alice, who is engaged in deliberations whose outcome is uniformly undetermined, is not doubly able, may be generalized: essentially the same argument shows that *no one* who is engaged in deliberations whose outcome is uniformly undetermined is doubly able.

The argument I have presented is one version of (the “third strand” of) the *Mind* argument. It is, at least to my mind, a powerful and persuasive argument. That is not to say that it persuades *me*. “History”, Gibbon tells us, “[is] little more than the register of the crimes, follies, and misfortunes of mankind.” I believe that the crimes and follies, if not in every case the misfortunes, of mankind are generally someone’s *fault* – that in most cases various people, or groups of people, are or have been to *blame* for each of them. I believe that if no one is or ever has been doubly able, then no crime or folly that history records is or ever was anyone’s fault. I believe (on the basis of the Consequence Argument) that if the outcome of a person’s deliberations on some occasion were determined, it follows that that person, on that occasion, was not doubly able. I believe, therefore, that there is something wrong with the *Mind* argument. But I am damned if I know what it is. I am *inclined to think* that (in the version of the argument that I have set out in the present Reply), the false step occurs somewhere in the reasoning by which I defended the second premise of the first-person argument ascribed to Alice:

If it is uniformly undetermined whether I shall decide to answer Yes, then I am not in a position to promise to answer Yes.

But I have no idea *what* is wrong with that reasoning. It looks good to me. But then the following argument looked good to Zeno (a *very* able philosopher):

At any given instant, an arrow in flight is not moving. (Look at a painting that shows an arrow in flight. The painting, of course, shows the arrow as it is at some instant. Well, *look* at it: the arrow is just hanging there motionless in the air, isn’t it?) But if at every moment an arrow in flight is not moving, then that arrow never moves. But if anything ever moves, arrows in flight move. Hence, nothing ever moves.⁵³

This argument is defective, and we now know where the defect lies. I would contend, however, that no Greek of the fifth century B.C. had the conceptual resources to discover what the defect in the argument was. (Or the fourth century. Aristotle, realizing that *something* had to be wrong with the argument, attempted to say where it went wrong⁵⁴. His attempt was an abject failure.) But I think that if a contemporary of Zeno’s had said something along the lines of

I don’t see anything wrong with that argument. Considered simply as a piece of reasoning, it seems unassailable. But, since its conclusion is obviously wrong, there has to be flaw in it – although I’m damned if I know what it is.

⁵³This is not *quite* Zeno’s arrow argument as Aristotle states it in *Physics*, 239b 30–33. Indulge me.

⁵⁴*Loc. cit.*

that would have been a perfectly reasonable response to the argument. And that response is precisely my response to the *Mind* Argument – and I believe that it is as reasonable as the response of my imaginary Greek to Zeno’s argument. I do not say that it is a philosophically satisfying response, for, of course, it is not. But it is no less reasonable for being unsatisfying.

12.7.2 *How Exactly Does Free Will Work?*

The authors rightly note that I have never attempted to answer the question, how “free will” exactly works? Or, since, as I have said, I have “grave reservations” concerning the technical vocabulary that usually figures in discussions of “the free-will problem” (including that phrase itself), I will frame the question this way:

An agent is trying to decide whether to do *X* or to do *Y*. At every moment in this process, the agent is doubly able. At every moment in this process the outcome of the agent’s deliberations is uniformly undetermined. In the event, the agent decides to do *X*. (It of course follows from the agent’s having been doubly able during the course of his or her deliberations that the agent was able to have decided to do *Y*.) How, exactly, would one go about constructing a plausible – or even a *possible* – causal model of the way in which the agent’s deliberations eventuated in a decision to do *X* that was consistent with these conditions?

The authors, in effect, defend the thesis that no such model is possible – or at least that no such model is possible that is consistent with known physics. I find no defect in their arguments. (I do not understand everything they say, but I understand enough of what they say to be willing to concede that they have presented an argument for this conclusion, or a conclusion essentially equivalent to it, and that I find no defect in it.) So to argue, in fact, is to present a version of the third strand of the *Mind* Argument. For, of course, deliberation or decision-making is a causal process. If, therefore, an agent reaches a decision under any conditions whatever – including, of course, the conditions set out in my statement of the question – there must *be* a causal model of the process that terminated in that decision, a model that is not only possible but *correct*. Therefore, if no causal model of the decision made by an agent who was doubly able at every moment during a uniformly undetermined process of deliberation is possible, then it is not possible for there to be an agent who is doubly able at every moment during a uniformly undetermined process of deliberation.

I am sorry to say that the only response I have to their arguments is “Zenonian”:

The many “crimes and follies of mankind” were (most of them, at least) someone’s *fault*. But if no one has ever been doubly able, none of them would ever have been anyone’s fault. The criminals and fools who were the authors of these crimes and follies were therefore sometimes doubly able. But the Consequence Argument is sound, and, therefore, the deliberations that led the criminals and fools of history to commit their crimes and follies – crimes and follies that were

their fault – were (mostly) uniformly undetermined. There is, therefore, a correct causal model of the decisions made by agents who are doubly able at every moment during uniformly undetermined processes of deliberation. (And I have no doubt that this model is consistent with the laws of nature – the laws of nature as God sees them, at any rate.) There is, therefore, a flaw in the arguments that purport to demonstrate that no such model is possible. I concede, however, that no one is able to provide even a *hint* as to what the structure of such a model might be, and no one has any idea what is wrong with the very persuasive arguments for its impossibility.

12.7.3 *The Problem of Free Will for Dualists*

The authors at several places suggest that that “the problem of free will” is a more difficult problem for physicalists than it is for dualists. In “The Mystery of Metaphysical Freedom” (van Inwagen 2008b), I defended the following thesis:

The problem of free will (I allow myself this phrase for the sake of concision) would arise in precisely the same form in a world in which there were agents but there was nothing physical at all – a world in which all agents were immaterial intellects (“angels,” let us call them), provided only that in that world there was such a thing as the passage of time and that the angels sometimes deliberated about future courses of action.

I stand by the arguments I presented for that thesis.

12.7.4 *Free Actions and Probabilities*

In “Free Will Remains a Mystery” (van Inwagen 2000b), I made use of the following intuition pump (my statement of the intuition pump employs the vocabulary of the present Reply):

Suppose Alice is deliberating about whether to lie or to tell the truth. Suppose the outcome of her deliberations is uniformly undetermined. Suppose she eventually decides to tell the truth. Let t be some moment during the course of her deliberations. Imagine that, after she has told the truth, God causes the universe to revert to precisely the state it was in at t and then allows the time-evolution of the universe to “proceed normally” till the moment at which Alice reaches a decision. “This time”, Alice might decide to lie – or, of course, she might decide to tell the truth. And if God causes the universe to revert to precisely the state it was in at t a very large number of times, we shall probably observe that Alice tells the truth in a certain definite proportion of the “replays” (and that she therefore lies in a certain definite proportion of the replays). As the number of replays increases,

then we shall probably⁵⁵ observe that the ratio of “truths” to “lies” converges on some number m/n – where, on each of these “occasions,” prior to her decision to lie or tell the truth, the probability that she would tell the truth on that occasion was $m/(m + n)$ and the probability that she would lie was $n/(m + n)$.

The authors doubt whether we should expect to observe such a convergence – whether we should expect (*almost certainly*) to observe the truth-lie ratio converging on some such number as $1/1$ or $3/1$ or $211/773$ as the number of replays increases without limit.

If I have understood their argument, it is that *if Alice has free will*, we cannot expect to observe convergence. They write:

But [as we have shown] such an exact probability distribution is not compatible with a free will, as the undetermined decision-making process would be ruled by chance. (Sect. 8.5)

But that was my *point* – or at least my point was that it is very plausible indeed to suppose that this is so. The quoted sentence is, indeed, simply a statement of the conclusion of the third strand of the *Mind Argument* (or one version of it). My contention was that the “exact probability distribution” is a consequence of Alice’s deliberations being undetermined at t (the “replay moment”).⁵⁶ And if the “exact probability distribution” is incompatible with Alice’s “free will” (or, as I should prefer to say, with her being doubly able while she is deliberating about whether to lie or to tell the truth), it follows that her having free will (her being doubly able) is incompatible with the outcome of her deliberations being uniformly undetermined.

Now why do I say that an “exact probability distribution” is a consequence of the outcome of Alice’s deliberations being undetermined at the moment t ? Well, consider the set of “possible futures” that confront the universe at t – futures in which there was no break or causal discontinuity in the time-evolution of the universe at t and in which the laws of nature (which are never violated – *except*, in our intuition pump, on the occasions on which God miraculously causes the universe to revert to the precise state it was in at t) continue to be what they had always been. If the universe is deterministic (if its time-evolution is governed by deterministic laws), there is only *one* such possible future: the actual future, the one that is determined in every detail by “nature and nature’s laws” – by the state of the universe at t and the laws that govern its time-evolution. And if the universe is indeterministic, there is more than one such future. If there is more than one such future, there are infinitely many – at least as many as there are real numbers (and possibly as many as there are continuous functions of a real variable). If probabilities are real numbers, the prob-

⁵⁵“Probably”: if we observe repeated tossings of a fair coin, we shall *probably* observe that the ratio of the number of “heads” to the number of “tails” converges to $1/1$ or 1 as the number of tosses increases. But if the coin is tossed 10,000 times, it *might* fall “heads” every time – or every time but once, or every time but twice ... But of course such outcomes are vastly improbable.

⁵⁶Note that if “There is an exact probability distribution” is a logical consequence of “The outcome of Alice’s deliberations is undetermined”, it must also be a logical consequence of “The outcome of Alice’s deliberations is undetermined and Alice has free will”: if q is a logical consequence of p , it is a logical consequence of the conjunction of p and any other proposition.

ability of each individual future is 0.⁵⁷ The (objective) probability that a given event will occur after t is simply the *proportion* of the futures that confront the universe at t in which that event occurs. For example, the probability that the (indeterministic) coin I am about to toss ten times will fall “heads” four times is the proportion of the possible futures in which the coin falls “heads” four times. (If the coin is fair, this proportion is easy to calculate: one simple, if inefficient, way to calculate it would be to write out all the 1024 ten-term sequences of H’s and T’s, count the “favorable outcomes” – the sequences that contain exactly four H’s –, and divide the number of favorable outcomes by 1024, that is, by the total number of outcomes, favorable and unfavorable.⁵⁸).

But what exactly is *meant* by, e.g., “the proportion of the possible futures in which the coin falls ‘heads’”? If the number of possible futures were finite, then that question would be easy to answer. If there are n possible futures and the coin falls “heads” in m of them, then the proportion of the possible futures in which it falls “heads” is m/n . But, as we have said, the number of possible futures is infinite – it is some transfinite number, and the number of possible futures in which it falls “heads” is that *same* transfinite number.

The resolution of this difficulty lies in the fact that some infinite sets have ‘sizes’ (“measures” is the technical term) that are not determined by their cardinality. Consider, for example, a straight line-segment in the Euclidean plane, which, following the mathematicians, we may identify with the set of its constituent points. Although it is entirely composed of points (as many points as there are real numbers), each of which is of 0 length, the line those points compose is of non-0 length – it is, let us say, 3 units long. But the line has infinitely many proper parts (proper subsets) of length 1. Each of these 1-unit-long parts is composed of the same number of 0-length parts as the 3-unit-long whole of which it is a part. Therefore, certain infinite sets of points, straight line segments, have measures – their lengths – that are not determined by their cardinality. Other infinite sets of points have other measures that are not determined by their cardinalities: area and volume, for example.

⁵⁷ Consider a simple universe that consists of an “eternal” ball that is at every moment either red or blue. The ball changes color indeterministically at one-second intervals with an equal probability of its next color being red or being blue: if the ball turns red, then, one second later, it will either remain red (probability 0.5) or change to blue (probability 0.5). A “possible future” for this universe at any moment is thus an infinite sequence that starts in some such way as this: RRBRBRBBBBRBR ... If probabilities are real numbers, then the probability of each of these infinite possible futures is 0 – owing to the fact that any real number greater than 0 is too large to be that probability: for any real number k greater than 0, there is a positive integer n such that every n -term sequence of R’s and B’s has a probability lower than k (the probability of every n -term sequence is $(\frac{1}{2})^n$ – and for any number k greater than 0, there is an n such that $(\frac{1}{2})^n$ is smaller than k), and, obviously, the probability of any infinite sequence must be lower than the probability of every finite sequence. What holds for this simple imaginary universe, holds *a fortiori* for the enormously complex real universe.

⁵⁸ The number of favorable outcomes is 210 ($10!/(4!(10-4)!)$). Hence, the proportion of futures in which the coin falls “heads” four times is about 0.205 – and the probability that the coin will fall heads four times is 0.205.

I make this supposition: that infinite sets of futures have measures formally similar to the generalizations of length and area and volume that figure in point-set topology.⁵⁹ (More, exactly *some* infinite sets of futures do. I shall expand on this statement presently.) To secure this formal similarity, I will suppose that the measures of the sets of futures that confront the universe at an instant t have the following formal features (modeled on the formal features of measures of sets of points): the empty set is of measure 0; a set containing one possible future is of measure 0; the whole set of futures that confront the universe at a given moment is of measure 1; if S is a finite or countably infinite set of sets of possible futures no two of which have a common member, then the measure of $\cup S$ is the sum of the measures of its members. We may then identify, e.g., the proportion of possible futures in which, on any given replay, Alice will tell the truth with the measure of the set of possible futures in which she tells the truth. If for example, the measure of that set is $\frac{1}{4}$, the proportion of futures in which she tells the truth is 0.25 or 25%. That is to say, the *probability* that she will tell the truth is 0.25.⁶⁰ And that must be the probability on *every* replay if it is the probability on *any*: since the state of universe is *precisely* identical at the start of each replay, and since the same indeterministic laws will govern the subsequent time-evolution of the universe, *everything* that is relevant to the probability of the outcome “truth” is identical in every replay. (Consider two tossed coins A and B, inhabitants of the same universe and thus subject to the same laws of nature, each of which is such that whether it falls “heads” or “tails” is – ontologically – undetermined. Suppose that A and B are *exact* duplicates, and that they are tossed in *exactly* same way in *identical* environments. The objective probability that A will fall “heads” *must* equal the objective probability that B will fall “heads,” for everything that is relevant to the probability that A will fall “heads” is precisely duplicated in the case of B.)

Now here is an application of the “law of large numbers”:

If the probability on each replay that Alice will tell the truth is p , then, as the number of replays increases without limit, the probability that the proportion of truth-outcomes will not be observed to converge to p approaches 0.

(For example, “If the probability on each replay that Alice will tell the truth is 0.7732, then, as the number of replays increases without limit, the probability that the proportion of truth-outcomes will not be observed to converge to 77.32% approaches 0.”) It does not follow from this statement that such a convergence *will* be observed. Suppose the probabilities of “truth” and “lie” are equal – both 0.5. There is no contradiction in saying that Alice will tell the truth on the original

⁵⁹Think of these measures this way. Dame Fortune throws a dart at a dartboard, each point of which represents a possible future. If two regions on the surface of the board of equal area, the dart is equally likely to hit either of them. The measure of a set of futures (a region on the surface of the board) is the proportion of the whole surface of the board occupied by that region: any set of points whose area is one-third the area of the board has a measure of $\frac{1}{3}$.

⁶⁰These probabilities are *objective* probabilities, not subjective probabilities or credences. The relation between objective and subjective probabilities is, of course, a central problem in the philosophy of probability.

12.8 Fatalism: Reply to M. Reinhart

The reader will note that there are several words that occur frequently in the author's essay that will not occur in this Reply. They are:

- “Fatalism”. Chapter II of *An Essay on Free Will* (van Inwagen 1983) was called “Fatalism.” In that chapter, the word was used for a philosophical thesis entirely unrelated to any of the issues discussed in the author's essay. I think it inadvisable to use the word in any other sense in the context of a discussion of the arguments of that book.
- “Determinism”, “indeterminism”, “compatibilism”, and “incompatibilism”. None of these terms is relevant to the topics addressed in chapter V. The topic of chapter V is succinctly stated in its title: “What Our Not Having Free Will Would Mean.” The four terms I have listed are relevant only to the question, “Why might one suppose that we do not have free will?” – a question that is addressed in every other chapter of the book but not in chapter V.

I will attempt to explain clearly what the arguments of Chapter V were supposed to accomplish. I was not in that chapter defending the thesis that we have free will. That is, to use a term I introduced in the previous Reply, I was not defending the thesis that we human beings are sometimes doubly able. (As I said in “The Problem of Free Will Revisited,” the term “free will” has acquired too many senses in the current philosophical literature for it to be any longer a useful term. I will use it only very sparingly in this Reply.) My project, rather, was twofold: (i) to examine the consequences of a person's *believing* that we human beings are never doubly able, and (ii) to attempt to lay out the logical implications of the proposition itself – the proposition that human beings are never doubly able. I will discuss these two projects in that order.

12.8.1 *The Consequences of a Person's Believing that We Human Beings Are Never Doubly Able*

Suppose Baron d'Holbach is deliberating between two courses of action – he is, let us say, trying to decide whether to write a pamphlet denouncing Voltaire's deism or to write a treatise against transubstantiation. (Both projects attract him, but he is not willing to take the time to do both.) In so deliberating, I maintain, Holbach *manifests in his behavior* two beliefs: a belief that he is able write the pamphlet against Voltaire and a belief that he is able to write the treatise against transubstantiation. (Just as, in the “suspicious husband” example, which the author mentions, that unhappy man manifests in his behavior a belief that his wife has, or may well have, a paramour – a belief he vehemently, and in some sense, sincerely, denies having.) For deliberation is a species of behavior: to deliberate is to *do* a certain thing – just

as to ride a bicycle or to give a lecture or to comfort a friend in distress are each of them cases of doing a certain thing.

But Holbach, although he deliberates like everyone else, has written,

The life of man is a line that nature orders him to inscribe on the surface of the earth, without his ever being able to swerve from it, even for an instant.⁶²

If Holbach believes that he is either going to write against Voltaire or against transubstantiation and that he is not going to do both these things, this statement commits him to believing that either he is not able to write against Voltaire or that he is not able to write against transubstantiation – for, if the “line” that is his future life does not include the composition of a treatise against transubstantiation, then to be able to compose a treatise against transubstantiation would be to be able to swerve from that line. Nevertheless, in virtue of deliberating about which of these things he will do, he manifests in his behavior both a belief that he is able to write against Voltaire and a belief that he is able to write against transubstantiation. He therefore has the following beliefs:

1. that he is able to write against Voltaire,
2. that he is able to write against transubstantiation,
3. that he is either unable to write against Voltaire or unable to write against transubstantiation.

His beliefs are, therefore, inconsistent. And so are the beliefs all philosophers who believe that their lives are “lines from which they are unable to swerve, even for an instant” – that is, all philosophers who believe that they are never doubly able.

Now what follows from this? Why nothing much, really. It certainly does not follow that deliberation is impossible for those who deny that they are ever doubly able. It is, after all, perfectly possible for one to have inconsistent beliefs. Nor does it follow that Holbach should either give up his belief that he is never doubly able or else refuse ever to deliberate between alternative courses of action. To refuse ever to deliberate would be madness. And, assuming that he does intend to continue to deliberate when deliberation is called for, why should this intention lead him to relinquish his belief that no one is ever doubly able – a belief the truth of which he regards as having been demonstrated by sound philosophical reasoning?

Should he cease to believe that no one is ever doubly able in order to avoid having contradictory beliefs? Well, one may ask, what is wrong with one’s having contradictory beliefs? Following Geach, I contend that the *only* thing wrong with one’s having contradictory beliefs is that it guarantees that one has at least one false belief. If Holbach accepts the thesis that his deliberating about whether to write against Voltaire or to write against transubstantiation logically implies that he believes that he is able to do both these things, and if he believes that one or the other of those two beliefs must be false (and if – being sane – he attempts to reach a decision about which course of action to pursue by rationally weighing the pros and cons of each),

⁶²Holbach ([1780] 1781, *Système de la nature* I xi, 161): “Notre vie est une ligne que la nature nous ordonne décrire sur la surface de la terre sans jamais pouvoir nous en écarter pour un instant.”

then, being capable of elementary reasoning, he will conclude that he is going to have a false belief even if he gives up his belief (which he regards as true) that no one is ever doubly able. Thus, giving up the true belief (for so he supposes it to be) that no one is ever doubly able would profit him nothing – and could be achieved only at the expense of ceasing to believe a philosophical truth of considerable importance. If it is pointed out to him that his beliefs are inconsistent, he should reply, “Ah yes, I have inconsistent beliefs. What of it?”

I thought it of *some* importance to point out that anyone who rejects the thesis that human beings are sometimes doubly able (and who deliberates – but everyone deliberates) must have inconsistent beliefs. It was never my purpose to contend that this fact has any significance that, as it were, goes beyond itself. If those who reject this thesis are willing to say, “Yes, I have inconsistent beliefs,” I do not regard that statement as constituting a *reductio ad absurdum* of their philosophical position.

12.8.2 *The Logical Implications of “No One Is Ever Doubly Able”*

In sections 5.3–5.9 (of van Inwagen 1983), I defended the thesis that if no one has or ever had free will, it follows that no one is, or ever has been, morally responsible for anything – or, as I should now prefer to say,

If no one is ever doubly able, it follows that nothing is or ever has been anyone’s fault – it follows that no one is or ever has been to blame for any state of affairs that has ever obtained or any event that has ever happened.

Thus, if the arguments of sections 5.3–5.9 are correct, and if no one is ever doubly able, then:

- The sad consequences of Professor Lustig’s unprofessional decision (see “The Problem of Free Will Revisited”, Chap. 1 in the present volume) were not his fault (or his student’s fault or anyone else’s fault).
- If a slum tenement burns to the ground and dozens of men, women, and children die horrible deaths in the fire, and if this has happened because the landlord bribed a fire inspector not to report the numerous fire-safety violations found in the building, neither the landlord nor the inspector (nor anyone else) is to blame for those deaths.
- No one is to blame for the Holocaust.
- The Atlantic slave trade was no one’s fault.

If the arguments of sections 5.3–5.9 are correct, and if no one is ever doubly able, then all these things were simply *misfortunes*. They are no more anyone’s fault than the death of the unfortunate V. Kamaraj, who was killed by a meteorite in India in 2016, or the many deaths that resulted from the eruption of Mount Vesuvius in A.D. 79 and the Lisbon earthquake of 1755 and the 2006 tsunami in the Indian Ocean.

This seems to me to be a fantastic position. If a chain of reasoning has led us to this position, we should conclude that that chain of reasoning was defective. We should regard it as most people would regard the reasoning that led Zeno to maintain that change is an illusion and the reasoning that led Descartes to maintain that non-human animals do not feel pain. That is, we should regard it as containing at least one mistake – at least one false premise or one invalid inference. And we should regard it as unsound whether we are able to identify that mistake or not.⁶³

12.9 Accountability: Reply to A. Strickmann and C. Weidemann

As I said in “The Problem of Free Will Revisited” (Chap. 1 in this book), I am distrustful (to use a very mild word) of the traditional technical vocabulary associated with the phrase “the problem of free will”. I will not in this Reply use any of the following words and phrases (items in that traditional vocabulary all) that figure prominently in the authors’ essay:

- (moral) responsibility
- (moral) accountability
- culpability
- (good) (moral) excuse for one’s acts.

(Nor will I use any of the close grammatical relations of these words and phrases – e.g., “responsible”.) In discussing situations of the kind to which philosophers tend to apply these terms, I will use only variants on phrases of the following forms (in various tenses):

- It is X’s fault that p
- E is X’s fault
- X is to blame for the fact that p
- X is to blame for E.

(Here “p” represents a declarative sentence and “E” represents a phrase that denotes an event or state of affairs.)

12.9.1 *The Strange Case of Henry T.*

The first question to be considered is a purely metaphysical question: the question whether there are in the story two persons who go by the name “Henry Turner”. Is the person the authors label “Henry-before-the-attack” a numerically different

⁶³The final paragraph of this Reply summarizes an argument presented not in chapter V but in chapter VI of *An Essay on Free Will* (van Inwagen 1983).

person from the person the authors label “Henry-after-the-attack”? I have in various places defended the position that the “two persons” thesis is metaphysically – even *logically* – incoherent. (Unless, that is, the metaphysician who accepts that thesis is either a proponent of the relativity of identity or a perdurantist. I cannot in this Reply discuss all the issues that have arisen in the vast literature on the metaphysics of identity across time in general and personal identity across time in particular. In the sequel, I am going simply to assume that only the strict numerical identity of the logic texts is relevant to questions of personal identity across time, and I am going simply to assume that perdurantism is a mistaken account of identity across time.) I refer the reader to two essays, “Materialism and the Psychological-Continuity Account of Personal Identity” (van Inwagen 1997), and “Temporal Parts and Identity across Time” (van Inwagen 2000a).

The story, therefore, contains only one person named “Henry Turner”. Now let us consider some bad state of affairs that is a consequence of some malfeasance of Henry’s in the days when he was a corrupt lawyer. Henry, let us say, bribed a crucial witness in a murder trial to lie under oath, and as a direct result, his client (who had indeed committed the murder) was acquitted, and Wilfrid, an innocent witness to the murder, was arrested, charged with the murder, and later convicted and sentenced to life in prison without the possibility of parole. Wilfrid, terrified by the prospect of spending the remainder of his life among the brutal inmates of a maximum-security prison, committed suicide in 1988, soon after his conviction. Suppose that, various conditions we could easily specify being satisfied, it was true *at the time* that Wilfrid’s death was Henry’s fault – it was true *in 1988* that Henry was to blame for Wilfrid’s death.⁶⁴

In 1991, however, Henry was, as the idiom has it, a changed man. Was Wilfrid’s death *still* Henry’s fault? Suppose someone (someone who knew all the facts of the case) said the following things to Henry in 1991:

Wilfrid’s miserable death in 1988 *was* your fault – *then*, in 1988. But *now*, in 1991, because you have become a changed man, it’s no longer your fault that Wilfrid died a miserable death. You *were*, in 1988, to blame for his death, but you’re no longer to blame for it. If you encounter Wilfrid’s widow (who has learned about the bribe, although she hasn’t any proof of it that would stand up in court), and she says to you, “*You*, Henry Turner! – it’s *your* fault that my husband was wrongly convicted and driven to suicide,” she’ll simply be wrong. She’ll only be saying that because she’s unaware of the transformation of your character. Or if she *is* aware of it, and, despite this knowledge, insists that her husband’s suicide was your fault, she’ll simply be philosophically confused.

Would that person have been right? Not in my view, certainly. The fact that Henry would not *now* (i.e., in 1991 – let us suppose, as a matter of expositional

⁶⁴Perhaps it is not *only* Henry’s fault. Perhaps it is not true that Henry *alone* is to blame for Wilfrid’s death. No doubt the perjured witness whom Henry bribed shares in the blame. Perhaps the actual murderer shares in the blame as well. Let this sort of qualification be understood in what follows.

convenience, that the present year is 1991) do things of the kind he once did (he would now – if he were still practicing law – recoil in horror from any suggestion that he should secure the acquittal of a man he *knew* to be guilty of murder by bribing a witness to commit perjury) does not change the fact that Wilfrid's death is *his fault*. After all, he *did* bribe the witness. It was *he* who did it; it was he who *did* it. He committed this act in the full knowledge of what he was doing. He was (we suppose this) *able* to have conducted his client's defense in full compliance with the law, but chose not to. His mental state at the time he bribed the witness exhibited *mens rea*. His act *did* have the consequence that a blameless man was, without warning, thrust into a situation in which it seemed to him that suicide was his only option.⁶⁵

Perhaps it will be instructive to compare the case of a wrong done with the case of a promise made. Suppose that in the 'old days', Henry kept a mistress, a woman named Wilma. (Such an arrangement would certainly have been consistent with Henry's other behavior at that stage of his life.) In 1988, Wilma, growing impatient with Henry's casual treatment of her, told Henry that if he did not divorce Sarah, and marry her, Wilma, she would go to Sarah and tell her everything. Thinking quickly, Henry told her that although he would not divorce Sarah, he would pay her, Wilma, \$100,000 in return for her silence – a promise he did not intend to keep and would in any case have been financially unable to keep. (His plan was to use the time the false promise gained him to gather materials that he could use to blackmail Wilma to secure her silence without paying her the promised money. But he was shot before he could put this plan into effect. Wilma – of course not knowing what the effect of his injury would be – did keep silent, intending to remind Henry of his promise when he had recovered.) The theses expressed in the above second-person passage seem to me to have no more merit than the theses expressed in this parallel passage:

You *did* promise to pay Wilma \$100,000 if she would keep silent – *then*, in 1988. (And, by the way, she *has* kept her part of the bargain.) But *now*, in 1991, because you have become a changed man, it's no longer the case that you have promised to pay Wilma \$100,000. If you encounter Wilma and she says to you, "You, Henry Turner! – you promised to pay me \$100,000 if I didn't tell your wife I was your mistress, and I haven't told her in all this time. Where's my money?" her statement that you made that promise to her will simply be wrong. She'll only be making that statement that because she's unaware of the transformation of your character. Or if she *is* aware of it, and, despite this knowledge, insists that you have promised to pay her \$100,000 in return for her silence, she'll simply be philosophically confused.⁶⁶

⁶⁵"In the story", Henry agrees with these judgments. The authors rightly point out that it does not follow from this that he is *right* to agree with them (see their n. 3, in Sect. 10.2), but his self-judgment is certainly worth noting.

⁶⁶This example is not intended to convince the reader that Henry is morally obliged to pay Wilma \$100,000. There can be all sorts of reasons why a person who has promised to do a certain thing is under no moral obligation to keep that promise. What I intend the example to establish is only this:

12.9.2 *The Strange Case of Walter B.*

Owing to no fault of his own, Walter B. was subject to very strong desires to commit immoral acts of many sorts. (The authors have stipulated “for the sake of argument” that these desires, however strong they may have been, were resistible.) The authors’ primary example of such an immoral act is Walter’s purchase of child pornography. There is a problem with this example from my point of view, however, because the consequences of an individual purchase of child pornography, while real and extremely serious, are hard to ‘trace’. The judicial opinion that the authors cite (in their note 5, Sect. 10.3) holds that an individual act of purchasing child pornography is “injurious to others” because it “supports a criminal industry” – and one might add, contributes causally, if only “statistically,” to the sexual abuse of the children whose abuse is recorded in the videos that are that industry’s principal product. It will be convenient to my argument to look at a more direct consequence of Walter’s act. Let us suppose that several of the videos Walter downloaded were viewed by his 12-year-old son, Alan, who, when Walter was at work, managed to hack into certain password-protected files on his father’s hard drive. (Alan, like most 12-year-old boys, was vastly more computer-literate than his father.) He did this because he suspected that the files were pornographic and he was eager to view them. (Alan assumed that they would depict adult sexual intercourse; he was scarcely aware that there *was* such a thing as child pornography.) As a result, Alan suffered severe psychological trauma and is still, years later, in therapy. Now let us ask: Was the trauma his son suffered Walter’s fault? Was he to blame for it? Well, of course he was: Walter performed a wrong act, an act he was able not to have performed, an act whose nature he understood, and any reasonably immediate consequence⁶⁷ of that wrong act (and Alan’s trauma was a sufficiently immediate consequence of his purchase of child pornography for this principle to apply) is Walter’s fault. He’s to blame for it⁶⁸ – just as, if he had sold a firearm illegally (to a shadowy figure who paid him cash in a dark alley), he would share in the blame for the death of anyone whom the purchaser later killed with it.

Henry – if he knows the facts of the case – cannot say to Wilma, “I did not promise to pay you \$100,000 if you kept silent.”

⁶⁷Imagine this case. A thief breaks into Walter’s house and steals his computer. Eventually, after several re-sales, the computer winds up in the hands of Desmond, who justifiably believes that he has purchased it legally. Through some operatic sequence of improbabilities, events having nothing to do with the files Walter downloaded, the authorities acquire the Gettier-style justified true belief that some of the files on Desmond’s new computer contain child pornography. A court order allows the police to impound the computer, one of their experts is able to open the protected files, and Desmond is charged with possession of child pornography. He is eventually found guilty of this charge, and an innocent man is sent to prison. The innocent Desmond’s imprisonment, I would judge, is not Walter B.’s fault: its causal connection with his purchase of child pornography is too remote. In the remainder of this Reply, when I speak of someone’s being to blame for the consequences of his or her acts, the qualification “reasonably immediate” is to be understood.

⁶⁸No doubt he’s not the *only* one who is to blame for it. No doubt the people who produced and sold the video the boy viewed also share in the blame. Cf. n. 64 in Sect. 12.9.1.

Is there anything more to say about this case? I do not think so, but the authors seem to. Consider Walter C., who is (once more, through no fault of his own) subject to the same strong but resistible desires to commit wrong acts as Walter B. Walter C., however, *does* resist those resistible desires and leads a wholly blameless life. Does Walter C. not deserve praise for having resisted these desires? It would not be *pleasant* to be Walter C., after all. The poor man's life is one long moral struggle. He has to engage in a daily battle with desires most of us are fortunate enough not to have. And, with terrible effort, he *wins* all those battles – battles that most of us are fortunate enough not to have to fight. Is he not a kind of moral hero? The authors, in effect, contend that he is. I have no quarrel with this contention. In fact, I accept it whole-heartedly. My hat, as the saying goes, is off to Walter C. What I do not see is the relevance of his case to any question of blame or fault or desert. The fact that Walter C. has played the very bad hand the universe dealt him and played it admirably does not change the fact that, in playing the same bad hand in a very different way, Walter B. acted wrongly. It does not imply that his son's psychological trauma is not Walter B.'s fault. It does not imply that Walter B. does not deserve both punishment under the law and the moral censure of his neighbors. Granted, Walter B. failed a very rarely administered test that Fate forced upon him and which those neighbors had the good fortune never to have to take. Granted, if he had passed the test, he would have been a praiseworthy man indeed. Nevertheless, he failed it and the bad consequences of the wrong things he did were his fault.

More or less the same point applies to Gyges. We rightly praise him for not misusing the powers that his ring granted him.⁶⁹ But if he *had* succumbed to the temptation to misuse his power to render himself invisible (it would have been a resistible temptation, for, as the authors tell his story, he did resist it), the bad consequences of his misuse of that power *would have been* his fault.

I think that the authors would disagree, although it is very hard to reconstruct their positions and arguments in terms I can accept. I might try this:

Van Inwagen accepts both the following theses.

- A and B are both tempted to do wrong and both resist these temptations. A's temptations are much stronger, much harder to resist, than B's. A, therefore, deserves greater praise for not doing wrong than does B.
- A and B are both tempted to do wrong and both give into these temptations. A's temptations are much stronger, much harder to resist, than B's. A and B are, nevertheless, both to blame for the bad consequences of their wrong acts.

But anyone who accepts the first should reject the second.

But I do not think that they would agree that this was an adequate statement of their argument. (In the unlikely event that they did accept this representation of their argument, my only reply would be: "I do not see why someone who accepts the first

⁶⁹As we praise Galadriel (from Tolkien's *Lord of the Rings*), who was tempted – *severely* tempted – by the powers of a ring and refused them: "I pass the test. I will diminish, and go into the West and remain Galadriel."

thesis is barred from accepting the second.”) One possibility is that they will say that this argument would be closer to their argument if “*equally to blame*” were substituted for “both to blame” in the second of the two theses. That is, they might contend that the right position is that although A and B are both to blame *to a certain degree* for the bad consequences of their wrong acts, B is to blame for those bad consequences *to a lesser degree* than A.

The new second thesis, the thesis produced by the proposed substitution, would not be *my* thesis, however. For I am not willing to concede that fault or blame comes in degrees. In my view, a certain bad event or state of affairs is either (e.g.) Alice’s fault or ... it is not. Having made this bald statement, I must qualify it in two ways.

First, fault and blame, like almost every concept we employ in conducting our lives are vague concepts, concepts that admit of borderline cases. To make one very obvious point, “reasonably immediate consequence” is certainly a vague concept, and there can therefore be bad consequences of a person’s wrong act such that there is no determinate answer to the question whether those consequences were “reasonably immediate” – and therefore no determinate answer to the question whether those consequences are the fault of the person who performed the act. This fact, I take it, is irrelevant to the question whether fault or blame comes in degrees.

Secondly, there are various idiomatic phrases that might seem to imply that fault or blame comes in degrees: “it’s not entirely my fault that ...” for example, or “she’s not wholly to blame for ...”. But I do not think that they really do imply that. To my ear, at any rate, the function of such phrases is simply to make the point that others share in the blame for the event or state of affairs under discussion. We might, for example, expand on the statement, “Sally was not wholly to blame for the accident” by saying: “Granted she was texting and was not properly attentive – but we must not forget that the driver of the car she struck cut in suddenly and sharply in front of her without even signaling that he was about to change lanes.”

I am, of course, aware that many philosophers who talk of “moral responsibility” and who suppose that moral responsibility is a more fundamental notion that underlies both praise and blame suppose that praise and blame exhibit some sort of symmetry, but *I* do not suppose this.⁷⁰ I would point out that we praise people for – among other

⁷⁰The authors point out in their note 9 (Sect. 10.3.4) that in the chapter on free will in *Metaphysics* (a textbook for undergraduates), I said in an aside (in this context: I was discussing various things that might be said against the thesis that human beings lack free will): If the lack of free will “rules out blame, it may well rule out praise on the same grounds.” I made that incidental remark only because “Lack of free will rules out praise” is a position that some philosophers have taken. (In this note, I understand “free will” in what one might call its classical sense: as the ability to act otherwise than one does.) I made no use of that thesis in the sequel and, indeed, said nothing further about praise. I would also point out that if the pope tells us that Jesus cannot be praised for being without sin because he was unable to sin and then goes on to state that Mary – who *was* able to sin, although she never did – is deserving of the highest praise for being without sin, he would not have affirmed two contradictory propositions. The first of these imaginary papal pronouncements is a consequence of the position that lack of free will rules out praise. That position has no relevance whatever to the question whether Mary – who did *not* lack free will in the matter of sinning – deserves praise for being without sin. And the cases of deserved moral praise the authors’ arguments appeal to are all cases in which someone was *able* to do wrong but did not.

things – their *actions* (“My hat is off to Walter C.”), their *characters* (“He is a model to us all”), their *abilities*, (“She is the most original philosopher working today”), and even their *appearances* (“What an astonishingly beautiful woman she is!”). We blame people for the *consequences* of their actions – or, it may be, of their inactions or failures to act (“It’s your fault your son is in therapy”; “You are to blame for my husband’s suicide”). Praise and blame are, therefore, not even directed at objects of the same sorts.⁷¹

In any event, there seems to me not to be even a *prima facie* case to be made for the following thesis:

We give higher praise to people who do the right thing when they are subject to abnormally strong temptation to do wrong than we do people who do the right thing when subject only to ‘ordinary’ or ‘normal’ temptation to do wrong. Reflection on this fact should lead us to say that people who are subject to abnormally strong temptation to do wrong and give in to this temptation are less to blame for the bad consequences of their wrong acts than are people who perform the same wrong acts when subject only to ‘ordinary’ or ‘normal’ temptation to do wrong.

Gyges, we have said, never misuses the power that his ring gives him, despite his knowledge that, if he chose to, he could enrich himself with impunity or harm those he disliked with no chance of being detected. (He’s tempted to misuse this power, but he resists the temptation.) Suppose that Archias has a similar ring and is tempted to use its powers to murder his father Pelopidas (with no possibility of anyone’s knowing that he has done this) in order to inherit the latter’s vast estates. Archias gives in to this (resistible) temptation and murders his father. (We know of the murder only because Archias confessed it on his deathbed.) What is wrong saying both the following things?

All praise to Gyges for never having misused the power of invisibility, despite the fact that he could have used that power to become rich by robbing his neighbors and to take vengeance on his enemies and to commit all manner of other crimes with no possibility of detection!

and

Many of the idle gossips who busy themselves with such matters have blamed Pelopidas’ wife for his untimely and mysterious death, and others one of his servants or concubines. But his son Archias’ deathbed confession has proved that the blame was Archias’ and his alone. One of the philosophers has said that Archias’ knowledge that he could murder his father with impunity was a *mitigating circumstance*, and that we should judge him less harshly than we should judge a man who had murdered his father for his estates and who had done so

⁷¹A qualification is needed. We may indeed blame people for their characters, their lack of ability, or their appearances – but *only if* we regard these items as consequences of their actions. But we may praise people for their characters (etc.) even if we believe that they are “unearned”, sheer gifts of fortune that are unconnected with anything their recipients have done.

knowing that there was a possibility that his crime would be detected. As Cicero has remarked, “There is nothing so absurd but some philosopher has said it”.

12.9.3 *Original Sin*

As one might guess from the second of the above imaginary conversations, I accept the following principle:

Suppose that A is subject to abnormally strong (but resistible) temptations to perform a certain wrong act, that A yields to these temptations, and that his performance of this act results in some bad thing. For the sake of concreteness, let us say that this bad thing is the death of Schmidt, an innocent man. Suppose, further that B is also tempted to perform this same wrong act and also yields to temptation and that B’s performance of this act also results in the death of an innocent man – in this case, Müller. Suppose that B’s temptations are of ‘ordinary’ or ‘normal’ strength. A is to blame for Schmidt’s death. B is to blame for Müller’s death. And as far as fault or blame are concerned, nothing more need be said. The abnormal strength of A’s desires was not a mitigating circumstance. If the wrong act in question is, let us say, “texting while driving”, the fact that various features of A’s psychological economy cause him to be very strongly tempted to text while driving does not mean that he should be judged less harshly than B (whose desire to text while driving is much weaker) for engaging in this dangerous activity. And, if both are (according to the laws of their country) found guilty of the misdemeanor “reckless endangerment”, the strength of A’s and B’s (resistible) desires to text while driving provide no reason for A’s receiving a lighter sentence than B.

It may be inferred from various statements of the authors that they find this principle highly counterintuitive. They speculate that my reasons for accepting “such a counterintuitive view” might be theological. But they are mistaken about the content of the “extended free-will defense”.⁷² They write:

However, this (almost) irresistible tendency to selfishness and evil provides us with no excuse. Sinners do not *deserve* divine indulgence (although God may forgive them by an act of grace): “It would have been just of God to leave human beings in the ruin they had made of themselves and their world.” (Sect. 10.3.6)

⁷²They are also mistaken about the meaning I give to the words “original sin”. In my use, at least, original sin is so-called because it is a part or aspect of all post-Fall human beings (or almost all: every Christian insists on one exception, and Roman Catholics on two) from and because of their *origins* – it was a part of them from the first moment of their existence, and it was conferred upon them by the very process that brought them into existence. “Original sin” does not refer to a certain pre-historical event – the event that brought about the Fall, the event that is described in mythical form in the bible in the story of the disobedience of Adam and Eve.

But, according to the extended free-will defense, the human tendency to selfishness and evil that is consequent on the Fall is not *almost* irresistible; it is irresistible, *simply* irresistible, irresistible without qualification. (That is to say, although each of us may be able to resist various particular temptations to do wrong, it is inevitable that we shall all do some wrong or other, that we shall eventually meet temptations we are unable to resist.) But God offers us a way out, a path into a condition in which we shall not invariably do evil – and shall eventually do no evil at all. It is offered to us: and, by a special Grace, God renders us able to both to accept and to reject the offer. (This is not *logically* all that different from the case of an alcoholic who finds the temptation to drink irresistible – but who has a free choice about whether to enter a therapeutic program whose procedures would eventually result in his being able to resist the temptation to drink.)

12.10 Evil: Reply to B. P. Göcke, C. Schneider and A. Sindermann

12.10.1 *Object and Meaning of “Defense”*

I want to begin by expressing my dissatisfaction with phrases like “van Inwagen’s defense of evil” and “van Inwagen’s defense of the existence of evil” and “van Inwagen’s defense of human suffering”. I express this dissatisfaction in the form of a parable.

A journalist, one Jack Hughes, has published a best-seller called *The Kennedy Assassination and the Secret Service: The Real Story*. The central thesis of the book is that the assassination would not have occurred but for gross dereliction of duty by the members of the US Secret Service who were charged with protecting President Kennedy. An historian, a Professor Konter, a specialist on the Kennedy administration, publishes an article called “The Unreal Story: The Bizarre Allegations of Jack Hughes” in the *New York Review of Books*. The burden of the article is that the conduct of the Secret Service on 22 November 1963, and in the weeks leading up to that day, was beyond reproach, and that Hughes’s arguments to the contrary are a tissue of innuendoes, half-truths, and erroneous history. An op-ed piece attacking “The Unreal Story” shortly thereafter appears in the *New York Times* (it later transpires that its author was a close friend of Hughes). The op-ed piece is entitled “Professor Konter’s Defense of the Assassination of President Kennedy”.

What I defended in *The Problem of Evil* (van Inwagen 2006) was not the existence of evil or human suffering but belief in God. And what I defended it *against* was the contention, widespread among philosophers, that the existence of vast amounts of truly horrendous evil (which cannot be rationally denied) entails that God does not exist. Here is a simple example of an argument of the sort that is commonly

supposed to demonstrate this entailment (a sort of minimal version of what I have called the global argument from evil):

- (1) There is a vast amount of truly horrendous evil in the world.
- (2) If there were a God, there would not be a vast amount of truly horrendous evil in the world.

Hence:

- (3) There is no God.

The argument is logically valid, and its first premise is obviously true. My main project in *The Problem of Evil* was an attempt to show that there is no reason to accept premise (2). (That is, my project was not to show that (2) was false or that there was reason to accept its denial, but simply to show that there was *no* reason to accept *it*. Thus, my project was to show that (2) should be regarded by agnostics as having the same epistemic status as “The number of Fraser firs in the Black Forest at the present moment is odd” has for all of us – as being a proposition that they have no reason to accept and no reason to reject.) The motivation for this project should be clear enough: if an argument has a premise that there is no reason to accept, then that argument does not constitute a reason to accept its conclusion.

My method in attempting to show that there was no reason to accept (2) was to tell a story according to which God exists and the world contains a vast amount of truly horrendous evil and which has this further feature (or so I hoped): Given the existence of God and the facts we know about the world, the *rest* of the story is true for all anyone knows. (If p entails the falsity of q , and if p is true for all anyone knows, then no one has any reason to accept q .) A story that has these two features I call a “defense”. (The description of my method in this paragraph is vastly oversimplified. It leaves out of account the “debate” context in which I imagined the defense presented and evaluated.)

The authors contend that my proposed defense (“the Expanded Free-Will Defense” or “EFW”) can easily be seen to be false, since God is by definition morally perfect⁷³, and the being called “God” in EFW behaves very badly indeed. I think it is fair to say that their charge against “God” is summed up in the following quotation:

Van Inwagen’s God, even if He existed, is not a morally perfect being, but rather quite an average bully that wants to achieve a certain goal even at the expense of the suffering of the innocent. (Sect. 11.4)

12.10.2 *Free Will in the Defense*

Before I attempt to answer this charge, I want to clear up some apparent confusions of the authors about the “free will” that figures in the Expanded Free-Will Defense (EFW). The authors say:

⁷³“Shall not the Judge of all the earth do right?” (Gen 18:25)

Van Inwagen's defense of human suffering [*sic!*] is a free will defense based on a libertarian conception of freedom of the will ... A person acts freely if and only if there are no sufficient physical or mental conditions that entail particular decisions to be made by the corresponding person, which is to say that the person herself and her will is absolutely morally responsible for the actions taken. (Sect. 11.3)

I have always insisted – well, *long* insisted – that there is no such thing as a libertarian conception of free will (see in particular van Inwagen 2008a). As my essay in this volume shows (Chap. 1), I have renounced the term “free will”, but when I did speak of free will, I mean this by it: For one to have free will is for one, when one is deliberating about alternative courses of action, sometimes to be able to perform each of them. That at least some human beings have free will is the “free-will thesis”. “Compatibilism” is the position that the free-will thesis is compatible with determinism, that is, with the proposition that at every moment the past and the laws of nature together determine a unique future. “Incompatibilism” is the position that the free-will thesis (the *very same* thesis that figures in the definition of compatibilism) is incompatible with determinism. “Libertarianism” is the conjunction of the free-will thesis and incompatibilism. It follows from these definitions that libertarians and compatibilists who accept the free-will thesis mean *exactly the same thing* by “free will” (or by “*freier Wille*” or by “*libre arbitre*” or by “*liber voluntas*”) – which I suppose means that they have the same conception of free will. At any rate, I do not see what else “have the same conception of free will” could mean. The difference between them is that libertarians think that this one thing that both they and the compatibilists call “free will” is incompatible with determinism, and compatibilists think that *it*, that very same thing, is compatible with determinism. (Incidentally, the free-will thesis does not entail that anyone is ever ‘absolutely’ – whatever that means – morally responsible⁷⁴ for anything, for moral responsibility, if one insists on using this term, involves other factors than freedom. And, of course, many philosophers believe that moral responsibility does not *require* freedom.)

I will add that one of the propositions included in EFW is the proposition that free will is incompatible with determinism. And why not? – For the Consequence Argument (whose conclusion is incompatibilism) is sufficiently convincing that it is reasonable to contend that incompatibilism is, if not true, then at least true for all anyone knows.

12.10.3 *God's Morality in the Defense*

I now turn to the authors' charges against the morality of the God of EFW. I will begin by quoting a passage from my essay, “The Magnitude, Duration, and Distribution of Evil: A Theodicy”. (The authors cite the essay but do not consider the passage.)

⁷⁴I use the term “morally responsible” because the authors have used it. Readers who consult “The Problem of Free Will Revisited” (Chap. 1 in this volume) will see that I have no use for this term, which has, in my view, become encrusted with conceptual confusions, which cling to it like barnacles on the hull of a decaying ship.

Consider the parable of the Prodigal Son. [...] Suppose the father of the Prodigal had foreseen the probable effects of his son's rash use of his patrimony, and had hired actors to represent themselves as gamblers and deliberately to lose substantial sums to the Prodigal; and suppose that he had further arranged for his agents to bribe prostitutes to tell the Prodigal that they had fallen in love with him and wanted to give him all their earnings (following which declaration they are to pass on to him monies provided by his father); and suppose that the father's agents, on his instructions, had followed the Prodigal about in secret to protect him from the dangers attendant on the night life of the ancient Middle East. What would have been the effects of this fatherly solicitude? Certainly the son would have continued to squander his substance indefinitely and with impunity. But here the word impunity must be understood in a rather superficial sense: for the son will be living a life of illusion (and that is a misfortune), and it is hard to see what could ever induce him to consider returning to his father (and I am inclined to think that that would also be a misfortune). (van Inwagen 1988, 173)

It is particularly puzzling to me that the authors do not consider this passage, since they have unfavorably compared the concern that God (in the story) shows for his creatures and a certain fictional mother's love for her son. I ask the reader to consider both the story the authors present in the section of their paper called "The Love of a Mother" (Sect. 11.3.2) and my hypothetical additions to the parable of the Prodigal [= wastrel] Son – and to consider the latter in the light of EFW. Let us first clear up some confusions about EFW that emerge in the authors' story of "Sarah" and "Steve".

God [...] knew all along that the human beings he created had [a] genetic disposition to abuse their free will. (Sect. 11.3.2)

It is not a part of EFW that human beings had a genetic disposition to abuse their free will. What they had was free will – the freedom to turn away from God and to follow, in the words of the Book of Common Prayer, the devices and desires of their own hearts. They had no genetic disposition to exercise their free will by turning away from God; what they had was a free choice between turning away from God and remaining united to God in bonds of love. (God's reason for giving them a free choice in this matter is clearly set out in EFW.) It is also a part of the story that the human genome (for good evolutionary reasons) contains genes whose relation to the behavior and spiritual condition of human beings is as follows:

If pre-rational creatures who have inherited these genes are miraculously raised to rationality and are initially in union with God, these genes will have no effect on their behavior or on their inner spiritual life as long as they continue to be in union with God; if they at some subsequent point separate themselves from God, then, although they will continue to be rational beings, this inherited genetic makeup will come into play and will have 'free reign', with the consequence that they will be subject to irresistible desires that lead them to commit acts that are selfish and cruel and wicked (they will, moreover, lose the power they had while in union with God to protect themselves from the random, destructive forces of nature).

In other words, one may, by a free choice, lose one's free will. (But we know that if we know that we have any free choices at all: if one freely chooses to become a

heroin addict, one freely abandons free will.) But these genes played no part in the decision of our paradisaal forbears to turn away from God.

[O]nce mankind started their rebellion against God, the God [of EFW] kicked them out of paradise. (Sect. 11.3.2)

No, no, no. Their “starting their rebellion against God” and “their no longer being in paradise” are two names for the same state of affairs, for being in Paradise and being in union with God are one and the same thing. One might as well say that, according to the parable of the Prodigal Son, once the Son had demanded his inheritance, his father kicked him out of his house.

[God] only warned human beings once or a few times concerning the consequences of their free decision. (Sect. 11.3.2)

So far as I can remember, EFW included nothing about what sort of warnings paradisaal human beings had given about the consequences of separating themselves from God. But in a similar story I told elsewhere I said this:

The creatures who committed the initial act of rebellion received sufficient warning that their act would lead to disaster. While they may have been unlike us in many ways, they were not children and were at least as intelligent as we; they fully understood the warning and the wisdom and authority of its Source. (van Inwagen 1988, 164)

I once saw the following notice posted in a zoo (posted prominently in several places along the length of a three-meter-high fence of iron bars):

MOTHER GRIZZLY BEAR (*Ursus arctos horribilis*) AND CUBS
WARNING: IF YOU CLIMB THIS FENCE AND ENTER THE ENCLOSURE,
THE MOTHER BEAR WILL ATTACK AND KILL YOU!⁷⁵

Imagine that some fool does exactly what he has been warned not to do and is killed by the mother bear. His friends complain bitterly that the zoo “only put up a few warning signs”.⁷⁶

But now it is time to turn to what seems to me to be the core of the authors’ moral critique of the God of EFW: that he uses pain and suffering to bully separated humanity into loving him.

This is simply not so. Consider the Prodigal’s father. The Prodigal, after he had squandered his inheritance, was in a very bad way. And reflection on his condition led him to “come to himself” and make a decision:

And when he had spent everything, a great famine arose in that country, and he began to be in want. So he went and joined himself to one of the citizens of that country, who sent him

⁷⁵When I saw this notice, I was reminded of another notice, this one (in Greek, Latin, and Hebrew) in the Herodian Temple: “No Gentile may enter beyond the dividing wall into the court around the Holy Place. Whoever is caught will have himself to blame for his subsequent death.”

⁷⁶Why *would* our paradisaal ancestors have ignored an absolutely clear warning of the consequences of rebellion by a being they knew to be omniscient? For some interesting speculations, see the C. S. Lewis’s theological romance *Perelandra*.

into his fields to feed swine. And he would gladly have fed on the pods that the swine ate; and no one gave him anything. But when he came to himself he said: "How many of my father's hired servants have bread enough and to spare, but I perish here with hunger! I will arise and go to my father, and I will say to him, 'Father, I have sinned against heaven and before you; I am no longer worthy to be called your son; treat me as one of your hired servants.'" And he arose and came to his father. But while he was yet at a distance, his father saw him and had compassion, and ran and embraced him and kissed him. And the son said to him: "Father, I have sinned against heaven and before you; I am no longer worthy to be called your son." But the father said to his servants: "Bring quickly the best robe, and put it on him; and put a ring on his hand, and shoes on his feet; and bring the fatted calf and kill it, and let us eat and make merry; for this my son was dead, and is alive again; he was lost, and is found." (Luke 15: 12–24)

Let us suppose that the father was aware of the sufferings his son had brought upon himself. And let us suppose that only two alternatives were open to him: to act as in my hypothetical modification of the story (to see to that his son never suffered by hiring actors to represent themselves as gamblers and pretend to lose large sums to him, etc.) and to do nothing – to do nothing in the hope that his son's inevitable sufferings (for we may suppose that the father knew that his son's attempt to set up on his own would in due course lead to the misery that it indeed did lead to) would cause him to take stock of his life and "come to himself". Would anyone really say that, by choosing not to engage in surreptitious manipulation of his son's circumstances, the father used the misery that hidden tinkering with his son's life would have prevented to bully the son into returning to him? Or that the father "wanted to achieve a certain goal even at the expense of the suffering of the innocent"?

But would an omnipotent God, facing the problem of reconciling a fallen and separated humanity to himself, really be faced with a set of alternative courses of action analogous to those facing the father of the Prodigal (in my elaboration of the story) – the father who must either do nothing in the hope that his son will, as the cliché has it, see the error of his ways, or, by behind-the-scenes manipulation, protect his son from the natural consequences of his folly? The answer is Yes. The situation the God of EFW finds himself in is *precisely* analogous to the situation the father finds himself in. That God finds himself in a situation analogous to that of the Prodigal's father is a part of the story, a part of EFW. That part of the story is told on pp. 87–89 of *The Problem of Evil* (van Inwagen 2006).

I should mention, finally, that the sufferings of the present age are not presented in EFW as a motive for loving God. God is not represented as allowing those sufferings so that human beings will try to love him in order to avoid further misery. Rather, God is represented as allowing suffering (as opposed to 'canceling' it by an endless series of ubiquitous miracles) so that human beings will realize that there is something *drastically, radically wrong*. If the sufferings are represented as a *motive* for anything, it is for this: to consider seriously the question whether there might be some way out of the drastically, radically wrong persisting condition that we know as "human history".

12.10.4 *Probability in the Defense*

I turn now from the authors' contention that the God of EFW is a bully to some arguments of the authors that involve probability. This is a difficult task for me because I do not understand these arguments. I find myself unable even to summarize them. I therefore direct the readers of this Reply to the part of the authors' text that I propose to reply to without understanding it. The relevant text occurs in Sect. 11.3.1. It begins with the words "Let us proceed step by step", and ends with the words "No morally perfect being could do or accept this". (It perhaps also includes the paragraph that follows. I am not sure whether that paragraph is a continuation of the argument that precedes it because, as I have said, I do not understand the argument that precedes it – and, absent an understanding of an argument, it can be difficult to determine what point in a text constitutes the completion of the presentation of that argument.) I ask the reader who has studied that section of the authors' text to consider the following two points.

1. Suppose I know that each person's decision about whether to become a philosopher (each person among those who make such a decision) is a free decision – and therefore, by my lights, an undetermined decision. Suppose Ludwig is trying to decide whether to become a philosopher or an aeronaut. Suppose I concede that even God cannot be certain which decision Ludwig will make. God does know the *probabilities*. That is, God knows lots of facts of this general sort: At t , a certain moment during the course of Ludwig's deliberations, the (objective) probability that he will decide to become a philosopher is 0.7 and that the probability that he will decide to become an aeronaut is 0.3. Now if *God* does not know what Ludwig is going to decide, then *a fortiori* I do not know what he will decide. And let us suppose that what is true of my epistemic condition vis-à-vis Ludwig is true of my epistemic condition vis-à-vis everyone who is trying to decide whether to become a philosopher. Still, I think I can be quite confident that *some* of these people will decide to become philosophers. I think I am entirely justified in believing that the proposition "*No one at all* will apply for graduate study in philosophy in the next decade" is of very low probability indeed. I am justified in treating it for practical purposes as I treat "If I toss a coin twenty times, it will fall 'heads' every time" or "The next time Carolyn Lynch plays bridge, she will be dealt a hand of 13 cards of the same suit". Similarly, I would say, the in-principle unpredictability of *individual* free human actions is perfectly consistent with God's knowing such propositions as the following ("the plan" refers to his 'rescue plan' for humanity):

- If the plan is put into effect, the probability that no one will be saved is 0.0000000000013.
- If the plan is put into effect, the probability that everyone will be saved is 0.0000000000017.
- If the plan is put into effect, the probability that at least 10% of humanity will be saved is 0.732.
- If the plan is put into effect, the probability that at least 90% of humanity will be saved is 0.081.

2. The captain of a certain ship is attempting to rescue as many as possible of the passengers of another ship, which has been struck by an enemy torpedo and is sinking. She knows that if she does nothing, then all those people will die. She has a certain rescue plan in mind, and it is the best plan she can think of – that is, in her judgment, for each number n , where n is less than or equal to the number of passengers on the sinking ship, the plan she has in mind yields a greater probability that at least n passengers will be rescued than does any feasible alternative plan she can think of. She knows (somehow) that her plan will *probably* lead to the rescue of about 30% of the passengers on the doomed ship – although there’s a real possibility that it will lead to the rescue of only 20% of them and a real possibility that it will lead to the rescue of 50% of them. It is in fact, *possible* that it will lead to the rescue of none of them – but the probability of that outcome is very low, about 0.01. And it is *possible* that it will lead to the rescue of all the passengers – but, again, the probability of that outcome is very low, about 0.01.

Surely there is nothing morally objectionable about putting that plan into effect? And it seems evident that if the plan involves *not* providing morphine to the many passengers who were gravely injured by the explosion of the torpedo, if it involves leaving them in great pain because giving them morphine would seriously reduce their capacity to follow the complicated instructions that the captain’s rescue plan involves, she is justified in not alleviating their pain?⁷⁷

References

- Austin J. L. 1961. A plea for excuses. In *Philosophical Papers*. Oxford: Clarendon Press. (Originally published 1956 in *Proceedings of the Aristotelian Society* 57: 1–30.)
- Brenner, Andrew. 2016. *A defense of mereological nihilism*. Dissertation submitted to the Graduate School of the University of Notre Dame in partial fulfillment of the requirements for the degree of Doctor of Philosophy, February 2016.
- Gellner, Ernest. 1979. The last pragmatist, or the behaviourist platonist. In *Spectacles and predicaments: Essays in social theory*, ed. Ernest Gellner, 199–208. Cambridge: Cambridge University Press. Originally published in *The Times Literary Supplement*, 1975.
- Holbach, Paul H. Th. d’. [1780]1781. *Système de la nature, ou des loix du monde physique et du monde morale*, “nouvelle édition”, London.

⁷⁷The text cited in the authors’ note 12 (Sect. 11.3.1) is not relevant to EFW. That text has to do with the following purely theological problem. According to Christian theology, God has promised that some people will be saved. But if there is a certain probability – very, very low, perhaps, but not 0 – that no one will be saved how can God be in a position to make such a promise? If my solution to this problem is incorrect, if indeed the problem has no solution, then all that follows is that God, having put “the plan” into effect, should not have promised that his putting the plan into effect would lead to some persons being saved. (He should at most have promised that *almost certainly* everyone will be saved.) It does *not* follow that there is anything wrong with his putting the plan into effect. Consider our sea captain. Putting *her* plan into effect is obviously the right thing for her to do. But – at least this thesis seems very plausible to me – given that there is a probability of 0.01 that the result of putting the plan into effect will be that *no one* will be rescued, she should not *promise* that someone will be rescued.

- Kripke, Saul A. 2013. *Reference and existence: The John Locke lectures*. Oxford: Oxford University Press.
- Lewis, David K. 1978. Truth in fiction. *American Philosophical Quarterly* 15: 37–46.
- . 1983. *Philosophical papers: Volume I*. New York: Oxford University Press.
- . 1986. *On the plurality of worlds*. Oxford: Blackwell.
- Lewis, David K., and Stephanie Lewis. 1970. Holes. *Australasian Journal of Philosophy* 48: 206–212.
- van Inwagen, Peter. 1977. Creatures of fiction. *American Philosophical Quarterly* 14: 299–308.
- . 1983. *An essay on free will*. Oxford: Clarendon Press.
- . 1988. The magnitude, duration, and distribution of evil: A theodicy. *Philosophical Topics* 16: 161–187.
- . 1990. *Material beings*. Ithaca: Cornell University Press.
- . 1997. Materialism and the psychological-continuity account of personal identity. *Philosophical Perspectives* 11: 305–319.
- . 2000a. Temporal parts and identity across time. *The Monist* 83: 437–459.
- . 2000b. Free will remains a mystery. *Philosophical Perspectives* 14: 1–19.
- . 2003. Existence, ontological commitment, and fictional entities. In *The Oxford handbook of metaphysics*, ed. Michael J. Loux and Dean W. Zimmerman, 131–157. Oxford: Oxford University Press.
- . 2004. A theory of properties. In *Oxford Studies in Metaphysics: Volume 1*, ed. Dean Zimmerman, 107–138. Oxford: Clarendon Press.
- . 2006. *The problem of evil*. Oxford: Oxford University Press.
- . 2008a. How to think about the problem of free will. *The Journal of Ethics* 12: 327–341.
- . 2008b. The mystery of metaphysical freedom. In *Metaphysics: The big questions*, ed. Peter van Inwagen and Dean W. Zimmerman, 456–465. Oxford: Blackwell.
- . 2009a. Listening to Clifford's ghost. In *Conceptions of philosophy: Royal Institute of Philosophy Supplement*, ed. Anthony O'Hear, vol. 65, 15–35. Cambridge: Cambridge University Press.
- . 2009b. Being, existence, and ontological commitment. In *Metametaphysics: New essays on the foundations of ontology*, ed. David J. Chalmers, David Manley, and Ryan Wasserman, 472–506. Oxford: Oxford University Press.
- . 2011. Relational vs. constituent ontologies. *Philosophical Perspectives* 25 (Metaphysics): 389–405.
- . 2014a. Inside and outside the ontology room. In *Existence: Essays in ontology*, 1–14. Cambridge: Cambridge University Press.
- . 2014b. Elementary or not [Review of Kripke, *Reference and Existence*]. *Times Literary Supplement*, March 14, 2014: 25.
- Wittgenstein, L. [1921] 1984. *Logisch-philosophische Abhandlung*. In *Werkausgabe*, vol. 1, Frankfurt am Main: Suhrkamp. Translated by David F. Pears and Brian F. McGuinness as *Tractatus Logico-Philosophicus*, London: Routledge 1961.